AD________________

Award Number: DAMD17-98-1-8042

TITLE: New Classes of Conditional Toxins as Therapeutic Agents Against Breast Cancer

PRINCIPAL INVESTIGATOR: Alexander Varshavsky, Ph.D.

CONTRACTING ORGANIZATION: California Institute of Technology
Pasadena, California 91125

REPORT DATE: April 2001

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20011005 007

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
April 2001

3. REPORT TYPE AND DATES COVERED
Final (1 Apr 98 - 31 Mar 01)

4. TITLE AND SUBTITLE
New Classes of Conditional Toxins as Therapeutic Agents Against Breast Cancer

5. FUNDING NUMBERS
DAMD17-98-1-8042

6. AUTHOR(S)
Alexander Varshavsky, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
California Institute of Technology
Pasadena, California 91125

E-Mail: avarsh@caltech.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This report contains colored photos

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

During the three-year duration of this grant ((DAMD17-98-1-8042, the Idea Grant), we focused on the following projects:

1) Development of signal-regulated, cleavage-mediated toxins. During this time, we completed the work on the CUP9-mediated, HIV protease-dependent conditional toxins. A paper describing these results has been finished (its completion was delayed by difficulties in some of the necessary control experiments). This paper will be submitted for publication in the Fall of 2001.

2) We also carried out, during the first year of support, an extensive study whose major aim was to construct better portable degrons (degradation signals) that can be used, in particular, for designing conditional toxins. This effort was entirely successful. A paper describing these results was published at the end of 1999 (Suzuki & Varshavsky. *Degradation signals in the lysine-asparagine sequence space.* **EMBO J.** 18:6017-6026, 1999). A reprint of the paper is enclosed.

3) In 1999, we also completed the construction of a bivalent inhibitor of the N-end rule pathway, a useful reagent for conditional-toxin projects, and published this new approach (Kwon & Varshavsky. *Bivalent inhibitor of the N-end rule pathway.* **J. Biol. Chem.** 274:18135-18139, 1999).

4) Other projects that encompass the subject of this grant have also been carried out. They were less successful, given a variety of technical difficulties that we encountered, at the Idea-Grant stage of the work. These projects are briefly described below.

14. SUBJECT TERMS
Breast Cancer, proteolysis/degradation signals/codominance/conditional toxins/cancer

15. NUMBER OF PAGES
220

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
Unclassified

18. SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

19. SECURITY CLASSIFICATION OF ABSTRACT
Unclassified

20. LIMITATION OF ABSTRACT
Unlimited

THIS PAGE LEFT BLANK
INTENTIONALLY

Table of Contents

Cover……………………………………………………………………

SF 298……………………………………………………………………

Table of Contents…………………………………………………… 4

Introduction………………………………………………………… 5

Body………………………………………………………………… 5

Key Research Accomplishments………………………………… 9

Reportable Outcomes……………………………………………… 9

Conclusions………………………………………………………… 9

List of Supported Personnel……………………………………… 10

Appendices………………………………………………………… 11

Final Report for the DOD Grant DAMD 17-98-1-8042

Idea Grant, April 1998-April 2001

Introduction

The objective of research supported by this Idea Grant was to attempt the development of new designs of conditional toxins that are sensitive to the presence of more than one molecular target, and to begin implementing these new reagents as therapeutic agents against breast cancer.

Body

During three years of support by the Idea Grant we carried out several projects, described below.

Signal-regulated, cleavage-mediated toxins

During the last two years, we essentially completed the work on the first example of a cleavage-mediated conditional toxin. A draft of the paper about this class of conditional toxins, will be submitted for publication in the Fall of 2001 (a preliminary preprint of this work was submitted with the year-2000 Progress Report (I. Davydov & A. Varshavsky. *Conditional toxin against yeast cells expressing HIV protease*). (We were delayed in finishing this work by unexpected difficulties with remaining control experiments.)

Construction of degradation signals in the lysine-asparagine sequence space

Among the problems to be solved on the way to multitarget conditional toxins is the paucity of strong, portable N-degrons that are free of drawbacks of the existing degrons of this class. These drawbacks include the tendency of proteins bearing such degrons to undergo endoproteolytic cleavages that sever the N-degron from the rest of the protein, and thereby contribute to the high background of assays with N-degron-bearing proteins. As a part of our ongoing effort to construct better degradation signals, including better degrons for conditional toxins, we developed a new approach, which makes it possible to identify and isolate specific degradation signals in the sequence space of just two amino acids, lysine and asparagine A paper describing these results was published in late 1999. (T. Suzuki & A. Varshavsky. *Degradation signals in the lysine-asparagine sequence space.* EMBO J. 18:6017-6026, 1999.) (reprint enclosed.)

Bivalent Inhibitor of Protein Degradation by the N-End Rule Pathway

In a project relevant to the theme of this grant, we have developed a specific bivalent inhibitor of the N-end rule pathway. The relevance of this advance stems from the fact the bivalent inhibitor (described below) improved the understanding of the targeting (substrate-binding) sites of N-recognin, the E3 component of the ubiquitin-dependent N-end rule pathway, which is at the center of our sitoxins and comtoxin projects. Eventually, the designs of conditional toxins should be able to involve any of the several ubiquitin-dependent proteolytic pathways; the current focus on the N-end rule pathway reflects not only the interest of this laboratory but also the fact that this pathway is particularly well understood mechanistically. Moreover, in staying within the N-end rule pathway, it is possible to use the portable, highly active N-degrons, an advantage that is not available with most of the other ubiquitin-dependent pathways. A paper describing these results was published in 1999. (Y. T. Kwon & A. Varshavsky. *Bivalent inhibitor of the N-end rule pathway.* J. Biol. Chem. 274:18135-18139, 1999.) (reprint enclosed.)

Construction and Use of the Library Expressing Random Fragments of the Green Fluorescent Protein (GFP) and Src Protein Kinase

The aim of this project was to develop a generally applicable method for interfering with the *in vivo* function of a target protein of interest through the expression of a specific fragment of the target that would bind to a conformationally immature (nascent) target protein and thereby either arrest or strongly retard its conformational maturation. As a result, the complex of the target and its self-peptide would be nonfunctional insofar as the target's function is concerned. In addition, the interfering peptide would be equipped with a strong N-terminal degradation signal (the N-degron, previously studied by this laboratory). A permanently or transiently stable complex between a target protein and its self-peptide bearing an N-degron would be targeted for processive degradation by the N-end rule pathway, one of the pathways of the ubiquitin system. Once developed, this method could be applied to effect selective functional inhibition of undesirable proteins such as, for example, overexpressed oncoproteins. This trans-degradation method would also become a part of the comtoxin approach described in the Idea Grant that supports these studies. To develop the trans-degradation technique, we chose Green Fluorescent

Protein (GFP) as a model target protein. (Our next, clinically relevant, target is Src, a protein tyrosine kinase and an proto-oncoprotein activated in a number of human cancers, including breast cancer.) We carried out the construction of a library of clones which expressed polypeptide fragments derived from random segments of GFP. These fragments were expressed in the yeast *S. cerevisiae* from the P_{GAL} promoter as linear fusions to ubiquitin. As shown previously by this laboratory, the N-terminal ubiquitin moiety of the fusion is cotranslationally cleaved off by ubiquitin-specific proteases, making it possible to expose any desired amino acid residue at the N-terminus of the resulting polypeptide fragment. In developing this approach, we proceeded in steps, at first constructing a ubiquitin fusion-based library of GFP fragments as such. The second-generation library should have the same fragments bearing the 40-residue N-terminal extension called Arg-$e^{\Delta K}$, which contains the above-described N-degron. The library expressing random GFP fragments was constructed using DNase I in the presence of Mn^{2+} ions to fragment the GFP cDNA, and specific oligonucleotide adapters to insert the resulting fragments into the expression vector. In 2000, we began the construction of analogous libraries with v-Src and c-Src trans-targeting, degron-containing reporter. The v-Src and c-Src cDNAs were obtained from Dr. D. Morgan (Univ. of California, Berkeley). They were tested by sequencing the 5'- and 3'-ends (~200 bp each) of the coding regions, and confirming that the resulting sequences matched those for v-Src and c-Src in the GenBank database. Restriction fragments containing the Src genes were isolated by agarose gel electrophoresis, and are being used to construct random-fragment libraries as described above. Taking advantage of the unique Nco I site at the 5'-end the v-Src cDNA, the v-Src coding region (as a filled-in Nco I-Xho I fragment) was inserted into the integration vector pIP123-CUPR. This placed the Src ORF under the control of the copper-inducible P_{CUP1} promoter. This part project ran into technical difficulty, primarily because this promoter was insufficiently tight (not low enough level of Src in the uninduced state). (In contrast to the GFP-based library, where the screen is being carried out by fluorescence-activated cell sorting (FACS), the readout and the screen with Src are based on the known toxicity of mammalian Src to the yeast *S. cerevisiae.*) Examination of the three-dimensional structures of Src family tyrosine kinases shows that a loop connecting the SH2 domain to the kinase domain is bound to the SH3 domain of Src and sandwiched between the SH3 domain and the N-terminal half of the kinase when Src is in the inactive state. The kinase-inhibiting interaction requires phosphorylation of Src at its C-terminal region, itself bound to the SH2 domain. This area is a promising candidate for a trans-targeting N-degron. Our aims in this set of projects were, first, to optimize the library-based interference approach described above, using GFP as both the target and reporter, then to expand this approach to the Src kinase, and to link an N-degron to the interfering fragments identified in both the GFP and Src library screen. This multistep strategy should result in a generally applicable method for metabolically destabilizing (and thereby functionally inhibiting) any protein of interest, including the targets that must be down-regulated for either arresting or selectively killing the breast adenocarcinoma cells.

Unfortunately, over the last year, this ambitious project ran into technical difficulties, which remain to be overcome.

Designing Asparagine-Lysine Degrons That Work in Mammalian Cells

We began to test the lysine-asparagine N-degrons of *S. cerevisiae* that were identified in the preceding work supported by this grant (Suzuki & Varshavsky. *Degradation signals in the lysine-asparagine sequence space.* **EMBO J.** 18:6017-6026, 1999) (reprint enclosed) in mammalian cells. Our aim is to determine whether any of the highly active yeast N-degrons of this class that have been identified in yeast are comparably powerful in mammals. If they are, several applications, including those that encompass conditional toxins will become possible right away. If these degrons are not as strong in the mammalian-cell setting, we will modify them, using techniques that have already proved successful in yeast, to produce a set of degrons that would confer on a mammalian protein reporter an *in vivo* half-life of 3 minutes or less. With degrons of this quality, one could design proteolysis-based conditional mutants in mice, and conditional toxins as well. This project is continuing.

Enclosed with this Final Report are the recently published papers by the lab (1999-2001; three sets), and three copies of the PI's CV. The Final Report describes projects that were supported, either partially or entirely, by the DOD Idea Grant in 1998-2001.

Key Research Accomplishments

- Development of the cleavage-regulated conditional toxin.
- Construction of degradation signals in the lysine-asparagine sequence space.
- Designing asparagine-lysine degrons that work in mammalian cells.
- Construction and use of random fragments of GFP and Src as sources of trans-targeting fragments. (This latter project ran into serious technical difficulties and remains uncompleted.)

Reportable Outcomes

- Published papers (they are cited in the Body of the Final Report).
- Paper in preparation that reports the development of cleavage-regulated conditional toxin.
- Construction and characterization of highly active degrons in the Lys-Asn sequence space.
- Although the approach that involved the construction of protein fragments-based trans-targeting library has not been successful, thus far, this idea, supported by the Idea Grant,. was a valuable experience in the library design, and will contribute to our further effort in this area.

No invention disclosures or patents based on the work supported by this Idea Grant have been, or will be filed.

Conclusions

Please see Key Research Accomplishments above. The statements there are the main conclusions of work supported by this Idea Grant.

Appendices

Reprints of the cited papers by the lab, other papers by the lab (1998-2001), and PI's CV.

List of Personnel supported by the Idea Grant DAMD17-98-18042 (1998-2001):

Dr. Ilia Davydov

Dr. Anna Kashina

Dr. Rong-gui Hu

May 2001

CURRICULUM VITAE

Name	Alexander J. Varshavsky
Date and Place of Birth	November 8, 1946, Moscow, Russia
Citizenship	U.S. citizen.
Address	Division of Biology, 147-75 California Institute of Technology 1200 East California Boulevard Pasadena, CA 91125
Telephone	626-395-3785 (office); 818-541-9791 (home)
Fax	626-440-9821 (office fax); 818-248-5245 (home fax)
Email	avarsh@caltech.edu

Academic Appointments and Education

1970	B.S. in Chemistry, Department of Chemistry, Moscow University, Moscow, Russia
1973	Ph.D. in Biochemistry, Institute of Molecular Biology, Moscow, Russia
1973-1976	Research Fellow, Institute of Molecular Biology, Moscow, Russia
1977-1980	Assistant Professor of Biology, Department of Biology, M. I. T., Cambridge, Massachusetts
1980-1986	Associate Professor of Biology, Department of Biology, M. I. T., Cambridge, Massachusetts
1986-1992	Professor of Biology, Department of Biology, M. I. T., Cambridge, Massachusetts
1992-present	Howard and Gwen Laurie Smits Professor of Cell Biology, Division of Biology, California Institute of Technology, Pasadena, California

Other Appointments

1983-1987	Member, Molecular Cytology Study Section, NIH
1993	Co-organizer, the 1993 Banbury Conference on the Ubiquitin System, Cold Spring Harbor Laboratory, NY.
2001	Fellow, International Institute for Advanced Studies, Koto, Japan

Professional Societies:

Fellow, American Academy of Arts and Sciences, 1987

Member, National Academy of Sciences, 1995

Fellow, American Academy of Microbiology, 2000

Member, American Philosophical Society, 2001

Awards:

Merit Award, National Institutes of Health, 1998

Novartis-Drew Award in Biomedical Science, Novartis, Inc. and Drew University, 1998

Gairdner International Award, Gairdner Foundation, Canada, 1999
(with A. Hershko)

Sloan Prize, General Motors Cancer Research Foundation, 2000
(with A. Hershko)

Lasker Award in Basic Medical Research, Albert and Mary Lasker Foundation, 2000
(with A. Hershko and A. Ciechanover)

Shubitz Prize in Cancer Research, University of Chicago, 2000

Hoppe-Seyler Award, Society for Biochemistry and Molecular Biology, Germany, 2000

Pasarow Award in Cancer Research, Pasarow Foundation, 2001

Merck Award, American Society for Biochemistry and Molecular Biology, 2001
(with A. Hershko)

Wolf Prize in Medicine, Wolf Foundation, Israel, 2001
(with A. Hershko)

Special Lectures (recent):

Honors Program Lecture, New York University School of Medicine, 2000
Gordon Lecture, Brandeis University, 2000
Harvey Society Lecture, Rockefeller University, 2001
Marker Lectures, Pennsylvania State University, 2001
Nelson Lecture, Yale University, 2001
Keynote Lecture, Symposium on Oncogenes and Growth Control, Salk Institute, 2001

Selected Publications (1968-present)

(grouped by the fields; numbered chronologically)

Chromosome Structure and Gene Expression

1. Varshavsky, A. Regulation of synthesis of genetic repressors in bacteria. **Mol. Biol.** (Russia) 2:13-20 (1968).

4. Ilyin, Y. V., Varshavsky, A., Mickelsaar, U. N. and Georgiev, G. P. Redistribution of proteins in mixtures of nucleoproteins, DNA and RNA. **Eur. J. Biochem**. 22:235-245 (1971).

5. Varshavsky, A. and Georgiev, G. P. Clustered arrangement of histones F2a1 and F3 in chromosomal deoxyribonucleoproteins. **Biochim. Biophys. Acta** 281:449-674 (1972).

9. Varshavsky, A. and Georgiev, G. P. Redistribution of histones during unfolding of chromosomal DNA. **Mol. Biol. Reports** 1:143-148 (1973).

12. Varshavsky, A. and Ilyin, Y. V. Salt treatment of chromatin induces redistribution of histones. **Biochim. Biophys. Acta** 340:207-217 (1974).

14. Ilyin, Y. V., Bayev, A. A. Jr., Zhuse, A. L. and Varshavsky, A. Histone-histone proximity in chromatin as revealed by imidoester crosslinking. **Mol. Biol. Reports** 1:343-348 (1974).

17. Varshavsky, A. and Bakayev, V. V. Nu-bodies and free DNA in chromatin lacking histone H1. **Mol. Biol. Reports** 2:209-217 (1975).

23. Varshavsky, A., Bakayev, V. V., Chumackov, P. M. and Georgiev, G. P. Minichromosome of simian virus 40: presence of histone H1. **Nucl. Acids Res**. 3:2101-2114 (1976).

27. Varshavsky, A. Structural and functional organization of eukaryotic chromosomes. **Biol. Zentralblatt** 95:301-316 (1976).

30. Bakayev, V. V., Bakayeva, T. G. and Varshavsky, A. Nucleosomes and subnucleosomes: heterogeneity and composition. **Cell** 11:619-630 (1977).

31. Varshavsky, A., Nedospasov, S. A., Bakayev, V. V., Bakayeva, T. G. and Georgiev, G. P. Histone-like proteins in the *E. coli* chromosome. **Nucl. Acids Res**. 4:2725-2745 (1977).

33. Varshavsky, A., Bakayev, V. V., Nedospasov, S. A. and Georgiev, G. P. On the structure of eukaryotic, prokaryotic and viral chromatin. **Cold Spring Harbor Symp. Quant. Biol**. 42:457-472 (1977).

37. Varshavsky, A., Sundin, O. and Bohn, M. SV40 viral minichromosome: preferential exposure of the origin of replication. **Nucl. Acids Res**. 5:3469-3478 (1978).

38. Varshavsky, A., Sundin, O . and Bohn, M. A 400 bp stretch of SV40 viral DNA that includes the origin of replication is exposed in SV40 minichromosomes. **Cell** 16:453-466(1979).

39. Sundin, O. and Varshavsky, A. Staphylococcal nuclease makes a single nonrandom cut in the SV40 viral minichromosome. **J. Mol. Biol.** 132:535-546 (1979).

45. Levinger, L. and Varshavsky, A. Drosophila heat shock proteins are associated with nuclease-resistant, high salt-resistant nuclear structures. **J. Cell Biol.** 90:793-796 (1981).

50. Barsoum, J., Levinger, L. and Varshavsky, A. On the chromatin structure of the amplified, transcriptionally active gene for dihydrofolate reductase in mouse cells. **J. Biol. Chem**. 257:5274-5282 (1982).

51. Levinger, L. and Varshavsky, A. Protein D1 preferentially binds AT-DNA and is a component of *Drosophila melanogaster* nucleosomes containing AT-rich satellite DNA. **Proc. Natl. Acad. Sci. USA** 79:7152-7156 (1982).

54. Wu, K., Strauss, F. and Varshavsky, A. Nucleosome arrangement in green monkey α-satellite chromatin. **J. Mol. Biol.** 170:93-117 (1983).

67. Barsoum, J. and Varshavsky, A. Preferential localization of variant nucleosomes near the 5'-end of the mouse dihydrofolate reductase gene. **J. Biol. Chem.** 260:7688-7697 (1985).

73. Solomon, M. J., Strauss, F. and Varshavsky, A. A mammalian HMG protein recognizes a stretch of six AT base pairs in duplex DNA. **Proc. Natl. Acad. Sci. USA** 83:1276-1289 (1986).

80. Peck, L. J., Millstein, L., Eversole-Cire, P., Gottesfeld, J. M. and Varshavsky, A. Transcriptionally inactive oocyte-type 5S RNA genes of *Xenopus laevis* are complexed with TFIIIA *in vitro*. **Mol. Cell. Biol.** 7:3503-3510 (1987).

91. Winter, E. and Varshavsky, A. A DNA-binding protein that recognizes oligo dA-oligo dT tracts. **EMBO J.** 8:1867-1877 (1989).

DNA Replication, Gene Amplification, Transporters & Drug Resistance

41. Sundin, O. and Varshavsky, A. Terminal stages of SV40 DNA replication proceed via multiply intertwined catenated dimers. **Cell** 21:103-114 (1980).

43. Varshavsky, A. On the possibility of metabolic control of replicon "misfiring": relationship to emergence of malignant phenotypes in mammalian cell lineages. **Proc. Natl. Acad. Sci. USA** 78:3673-3677 (1981).

44. Varshavsky, A. Phorbol ester dramatically increases incidence of methotrexate-resistant cells: possible mechanisms and relevance to tumor promotion. **Cell** 25:561-572 (1981).

46. Sundin, O. and Varshavsky, A. Arrest of segregation leads to accumulation of highly intertwined catenated dimers: dissection of the final stages of SV40 DNA replication. **Cell** 25:659-669 (1981).

55. Barsoum, J. and Varshavsky, A. Mitogenic hormones and tumor promoters greatly increase the incidence of cells bearing amplified dihydrofolate reductase genes. **Proc. Natl. Acad. Sci. USA** 80:5330-5334 (1983).

56. Varshavsky, A. Diadenosine 5', 5''',-P^1,P^4-tetraphosphate: a pleiotropically acting alarmone? **Cell** 34:711-712 (1983).

58. Varshavsky, A. Do stalled replication forks synthesize a specific alarmone? **J. Theoret. Biol.** 105:707-714, 1983.

59. Snapka, R. and Varshavsky, A. Loss of unstably amplified dihydrofolate reductase genes from mouse cells is accelerated by hydroxyurea. **Proc. Natl. Acad. Sci. USA** 80:7533-7537 (1983).

63. Roninson, I., Abelson, H. T., Housman, D. E., Howell, N. and Varshavsky, A. Amplification of specific DNA sequences correlates with multidrug resistance in Chinese hamster cells. **Nature** 309:626-628 (1984).

71. Ciccarelli, R. B., Solomon, J. J., Varshavsky, A. and Lippard, S. J. *In vivo* effects of cis- and trans diaminedichloroplatinum (II) on SV40 chromosomes: differential repair, DNA-protein crosslinking, and inhibition of replication. **Biochemistry** 24:7533-7540 (1985).

72. Gros, P., Croop, J., Roninson, I., Varshavsky, A. and Housman, D. E. Isolation and characterization of DNA sequences amplified in multidrug-resistant hamster cells. **Proc. Natl. Acad. Sci. USA** 83:337-341 (1986).

81. Solomon, M. J. and Varshavsky, A. A nuclease-hypersensitive region forms *de novo* after chromosome replication. **Mol. Cell. Biol.** 7:3822-3825 (1987).

93. McGrath, J. P. and Varshavsky, A. The yeast *STE6* gene encodes a homolog of the mammalian multidrug resistance P-glycoprotein. **Nature** 340:400-404 (1989).

New Methods

16. Bakayev, V. V., Melnickov, A. A., Osicka, V. A. and Varshavsky, A. Isolation and characterization of chromatin subunits. **Nucl. Acids Res.** 2:1401-1419 (1975). *(Low ionic strength electrophoretic technique for separation of DNA-protein complexes.)*

22. Varshavsky, A., Bakayev, V. V. and Georgiev, G. P. Heterogeneity of chromatin subunits and location of histone H1. **Nucl. Acids Res.** 3:477-492 (1976). *(Fractionation of nucleosomes by the low ionic strength electrophoresis, the forerunner of gel shift assay.)*

42. Levinger, L., Barsoum, J. and Varshavsky, A. Two-dimensional hybridization mapping of nucleosomes. **J. Mol. Biol.** 146:287-304 (1981).

49. Boyce, F., Sundin, O., Barsoum, J. and Varshavsky, A. New way to isolate SV40 viral minichromosomes: use of a thiol-specific reagent. **J. Virol.** 42:292-296 (1982).

64. Strauss, F. and Varshavsky, A. A protein binds to a satellite DNA repeat at three sites which would be brought into proximity by DNA folding in the nucleosome. **Cell** 37:889-901 (1984). *(First use of gel shift assay for the detection of specific DNA-binding proteins in cell extracts.)*

69. Solomon, M. J. and Varshavsky, A. Formaldehyde-mediated DNA-protein crosslinking: a probe for *in vivo* chromatin structures. **Proc. Nat. Acad. Sci. USA** 82:6470-6474 (1985).

74. Swerdlow, P. S., Finley, D. and Varshavsky, A. Enhancement of immunoblot sensitivity by heating of hydrated filters. **Analyt. Biochem.** 156:147-153 (1986).

76. Snapka, R. M., Kwok, K., Bernard, J. A., Harling, O. and Varshavsky, A. Post-separation detection of nucleic acids and proteins by neutron activation. **Proc. Natl. Acad. Sci. USA** 83:9320-9324 (1986).

82. Varshavsky, A. An electrophoretic assay for DNA-binding proteins. **Meth. Enzymol.** 151:551-565 (1987).

83. Bartel, B. and Varshavsky, A. Hypersensitivity to heavy water: a new conditional phenotype. **Cell** 52:935-941 (1988). *(An alternative to temperature-sensitive conditional mutants.)*

86. Solomon, M. J., Larsen, P. L. and Varshavsky, A. Mapping protein-DNA interactions *in vivo* with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. **Cell** 53:937-947 (1988). *(A technique for localizing specific DNA-bound proteins to specific DNA sequences through in vivo crosslinking, DNA fragmentation, immunoprecipitation, and DNA hybridization. This method is called the CHIP (chromatin immunoprecipitation) assay.)*

112. Dohmen, R. J., Wu, P. P. and Varshavsky, A. Heat-inducible degron: a method for constructing temperature-sensitive mutants. **Science** 263:1273-1276 (1994).

113. Johnsson, N. and Varshavsky, A. Ubiquitin-assisted dissection of protein transport across membranes. **EMBO J.** 13:2686-2698 (1994).

115. Johnsson, N. and Varshavsky, A. Split ubiquitin as a sensor of protein interactions *in vivo.* **Proc. Natl. Acad. Sci. USA** 91:10340-10344 (1994).

123. Lévy, F., Johnsson, N., Rümenapf, T. and Varshavsky, A. Using ubiquitin to follow the metabolic fate of a protein. **Proc. Natl. Acad. Sci. USA** 93:4907-4912 (1996). *(A method for producing equimolar amounts of a reference and a test protein in vivo.)*

127. Johnson, N. and Varshavsky, A. Split ubiquitin: a sensor of protein interactions *in vivo.* In: **The Yeast Two-Hybrid System** (P. L. Bartel and S. Fields, eds.), pp. 316-332, Oxford University Press, N. Y. (1997).

137. Dünnwald, M., Varshavsky, A. and Johnsson, N. Detection of transient *in vivo* interactions between substrate and transporter during protein translocation into the endoplasmic reticulum. **Mol. Biol. Cell** 10:329-344 (1999). *(A method, based on the split-ubiquitin sensor, for detection of transient protein-protein interactions.)*

147. Turner, G. C. and Varshavsky, A. Detecting and measuring cotranslational protein degradation *in vivo.* **Science** 289:2117-2120 (2000).

148. Varshavsky, A. Ubiquitin fusion technique and its descendants. **Meth. Enzymology** 327:578-593 (2000).

Multitarget Compounds

117. Varshavsky, A. Codominance and toxins: a path to drugs of nearly unlimited selectivity. **Proc. Natl. Acad. Sci. USA** 92:3663-3667 (1995).

132. Varshavsky, A. Codominant interference, antieffectors, and multitarget drugs. **Proc. Natl. Acad. Sci. USA** 95:2094-2099 (1998).

The Ubiquitin System and Intracellular Proteolysis

40. Levinger, L. and Varshavsky, A. Separation of nucleosomes containing and lacking ubiquitin-H2A semihistone. **Proc. Natl. Acad. Sci. USA** 77:3244-3248 (1980).

48. Levinger, L. and Varshavsky, A. Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes in the *Drosophila* genome. **Cell** 28:375-386 (1982).

53. Swerdlow, P. and Varshavsky, A. Affinity of HMG17 for a nucleosome is not influenced by the presence of ubiquitin-H2A semihistone but depends on DNA fragment size. **Nucl. Acids. Res.** 11:387-401 (1983).

61. Finley, D., Ciechanover, A. and Varshavsky, A. Thermolability of ubiquitin-activating enzyme from the mammalian cell cycle mutant ts85. **Cell** 37:43-55 (1984).

62. Ciechanover, A., Finley, D. and Varshavsky, A. Ubiquitin dependence of selective protein degradation demonstrated in the mammalian cell cycle mutant ts85. **Cell** 37:57-66 (1984).

66. Özkaynak, E., Finley, D. and Varshavsky, A. The yeast ubiquitin gene: head-to-tail repeats encoding a polyubiquitin precursor protein. **Nature** 312:663-666 (1984).

70. Finley, D. and Varshavsky, A. The ubiquitin system: functions and mechanisms. **Trends Biochem. Sci.** 10:343-346 (1985).

75. Bachmair, A., Finley, D. and Varshavsky, A. *In vivo* half-life of a protein is a function of its N-terminal residue. **Science** 234:179-186 (1986).

77. Özkaynak, E., Finley, D., Solomon, M. J. and Varshavsky, A. The yeast ubiquitin genes: a family of natural gene fusions. **EMBO J**. 6:1429-1440 (1987).

78. Finley, D., Özkaynak, E. and Varshavsky, A. The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation and other stresses. **Cell** 48:1035-1046 (1987).

79. Jentsch, S., McGrath, J. P. and Varshavsky, A. The yeast DNA repair gene *RAD6* encodes a ubiquitin-conjugating enzyme. **Nature** 329:131-134 (1987).

84. Finley, D., Özkaynak, E., Jentsch, S., McGrath, J. P., Bartel, B., Pazin, M., Snapka, R. M. and Varshavsky, A. Molecular genetics of the ubiquitin system. In: **Ubiquitin** (M. Rechsteiner, ed.), pp. 39-75, Plenum Press, N. Y.(1988).

85. Varshavsky, A., Bachmair, A., Finley, D., Wünning, I. and Gonda, D. The N-end rule of selective protein turnover: mechanistic aspects and functional implications. In: **Ubiquitin** (M. Rechsteiner, ed.), pp. 287-324, Plenum Press, N. Y. (1988).

87. Goebl, M. G., Yochem, J., Jentsch, S., McGrath, J. P., Varshavsky, A. and Byers, B. The yeast cell cycle gene *CDC34* encodes a ubiquitin-conjugating enzyme. **Science** 241:1331-1335 (1988).

88. Bachmair, A. and Varshavsky, A. The degradation signal in a short-lived protein. **Cell** 56:1019-1032 (1989).

89. Chau, V., Tobias, J. W., Bachmair, A., Mariott, D., Ecker, D., Gonda, D. K., and Varshavsky, A. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. **Science** 243:1576-1583 (1989).

90. Finley, D., Bartel, B. and Varshavsky, A. The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. **Nature** 338:394-401 (1989).

92. Gonda, D. K., Bachmair, A., Wünning, I., Tobias, J. W., Lane, W. S. and Varshavsky, A. Universality and structure of the N-end rule. **J. Biol. Chem.** 264:16700-16712 (1989).

94. Balzi, E., Choder, M., Chen, W., Varshavsky, A. and Goffeau, A. Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. **J. Biol. Chem.** 265:7464-7471 (1990).

95. Hochstrasser, M. and Varshavsky, A. *In vivo* degradation of a transcriptional regulator: the yeast α2 repressor. **Cell** 61:697-708 (1990).

96. Johnson, E. S., Gonda, D. K. and Varshavsky, A. *Cis-trans* recognition and subunit-specific degradation of short-lived proteins. **Nature** 346:287-291 (1990).

97. Bartel, B., Wünning, I. and Varshavsky, A. The recognition component of the N-end rule pathway. **EMBO J**. 9:3179-3189 (1990).

98. Baker, R. T. and Varshavsky, A. Inhibition of the N-end rule pathway in living cells. **Proc. Natl. Acad. Sci. USA** 88:1090-1094 (1991).

99. Varshavsky, A. Naming a targeting signal. **Cell** 64:13-15 (1991).

100. Hochstrasser, M., Ellison, M. J., Chau, V. and Varshavsky, A. The short-lived Matα2 transcriptional regulator is ubiquitinated *in vivo*. **Proc. Nat. Acad. Sci. USA** 88:4606-4610 (1991).

101. Tobias, J. W. and Varshavsky, A. Cloning and functional analysis of the ubiquitin-specific protease gene *UBP1* of *S. cerevisiae*. **J. Biol. Chem**. 266:12021-12028 (1991).

102. Dohmen, R. J., Madura, K., Bartel, B. and Varshavsky, A. The N-end rule is mediated by the Ubc2 (Rad6) ubiquitin-conjugating enzyme. **Proc. Natl. Acad. Sci. USA** 88:7351-7355 (1991).

103. McGrath, J. P., Jentsch, S. and Varshavsky, A. *UBA1*: an essential yeast gene encoding ubiquitin-activating enzyme. **EMBO J**. 10:227-237 (1991).

104. Tobias, J. W., Shrader, T. E., Rocap, G. and Varshavsky, A. The N-end rule in bacteria. **Science** 254:1374-1377 (1991).

105. Johnson, E. S., Bartel, B., Seufert, W. and Varshavsky, A. Ubiquitin as a degradation signal. **EMBO J**. 11:497-505 (1992).

106. Ota, I. and Varshavsky, A. A gene encoding a putative tyrosine phosphatase suppresses lethality of an N-end rule-dependent mutant. **Proc. Natl. Acad. Sci. USA** 89:2355-2359 (1992).

107. Baker, R. T., Tobias, J. W. and Varshavsky, A. Ubiquitin-specific proteases of *S. cerevisiae:* cloning of *UBP2* and *UBP3*, and functional analysis of the *UBP* gene family. **J. Biol. Chem**. 267:23363-23375 (1992).

108. Varshavsky, A. The N-end rule. **Cell** 69:725-735 (1992).

109. Shrader, T. E., Tobias, J. W. and Varshavsky, A. The N-end rule in *Escherichia coli*: cloning and analysis of the leucyl, phenylalanyl-tRNA-protein transferase gene *aat*. **J. Bact**. 175:4364-4374 (1993).

110. Madura, K., Dohmen, R. J. and Varshavsky, A. N-recognin/Ubc2 interactions in the N-end rule pathway. **J. Biol. Chem** 268:12046-12054 (1993).

111. Ota, I. M. and Varshavsky, A. A yeast protein similar to bacterial two-component regulators. **Science** 262:566-569 (1993).

114. Madura, K. and Varshavsky, A. Degradation of Gα by the N-end rule pathway. **Science** 265:1454-1458 (1994).

116. Johnston, J. A., Johnson, E. S., Waller, P. and Varshavsky, A. Methotrexate inhibits proteolysis of dihydrofolate reductase by the N-end rule pathway. **J. Biol. Chem.** 270:8172-8178 (1995).

118. Baker, R. T. and Varshavsky, A. N-terminal amidase: a new enzyme and component of a targeting complex in the N-end rule pathway. **J. Biol. Chem.** 270:12065-12074 (1995).

119. Varshavsky, A. The world of ubiquitin. **Engineering & Science** 58:26-36 (1995).

120. Johnson, E. S., Ma, P. C. M., Ota, I. M. and Varshavsky, A. A proteolytic pathway that recognizes ubiquitin as a degradation signal. **J. Biol. Chem**. 270:17442-17456 (1995).

121. Dohmen, R. J., Stappen, R., McGrath, J. P., Forrová, H., Kolarov, J., Goffeau, A. and Varshavsky, A. An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. **J. Biol. Chem**. 270:18099-18109 (1995).

122. Varshavsky, A. The N-end rule. **Cold Spring Harbor Symp. Quant Biol**. 60:461-478 (1996).

124. Ghislain, M., Dohmen, R. J., Lévy, F., and Varshavsky, A. Cdc48p interacts with Ufd3p, a WD-repeat protein required for ubiquitin-dependent proteolysis in *Saccharomyces cerevisiae*. **EMBO J.** 15:4884-4899 (1996).

125. Varshavsky, A. The N-end rule: functions, mysteries, uses. **Proc. Natl. Acad. Sci. USA** 93:12142-12149 (1996).

126. Grigoryev, S., Stewart, A. E., Kwon, Y. T., Arfin, S. M., Bradshaw, R. A., Jenkins, N., Copeland, N. G. and Varshavsky, A. A mouse amidase specific for N-terminal asparagine: the gene, the enzyme, and their function in the N-end rule pathway. **J. Biol. Chem.** 271:28521-28532 (1996).

128. Varshavsky, A. The N-end rule pathway of protein degradation. **Genes Cells** 2:13-29 (1997).

129. Varshavsky, A. The ubiquitin system. **Trends Biochem. Sci**. 22:383-387 (1997).

130. Varshavsky, A., Byrd, C., Davydov, I. V., Dohmen, R. J., Ghislain, M., Gonzalez, M., Grigoryev, S., Johnson, E. S., Johnsson, N., Johnston, J. A., Kwon, Y. T., Lévy, F., Lomovskaya, O., Madura, K., Rümenapf, T., Shrader, T. E., Suzuki, T., Turner, G. and Webster, A. The N-end rule pathway. In **Ubiquitin and the Biology of the Cell** (D. Finley and J.-M. Peters, eds), Plenum Press, NY, pp. 232-278 (1998).

131. Byrd, C. Turner, G. and Varshavsky, A. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. **EMBO J.** 17:269-277 (1998).

133. Davydov, I. V., Patra, D. and Varshavsky, A. The N-end rule pathway in *Xenopus* egg extracts. **Arch. Biochem. Biophys.** 357:317-325 (1998).

134. Kwon, Y. T., Reiss, Y., Fried, V. A., Hershko, A., Yoon, J. K., Gonda, D. K., Sangan, P., Copeland, N. G., Jenkins, N. A. and Varshavsky, A. The mouse and human genes encoding the recognition component of the N-end rule pathway. **Proc. Natl. Acad. Sci. USA** 95:7898-7903 (1998).

135. Ramos, P. C., Höckendorff, J., Johnson, E. S., Varshavsky, A. and Dohmen, R. J. Ump1p is required for proper maturation of the 20S proteasome and becomes its substrate upon completion of the assembly. **Cell** 92:489-499 (1998).

136. Lévy, F., Johnston, J. A. and Varshavsky, A. Analysis of a conditional degradation signal in yeast and mammalian cells. **Eur. J. Biochem.** 259:244-252 (1999).

138. Kwon, Y. T., Kashina, A. S. and Varshavsky, A. Alternative splicing results in differential expression, activity and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway. **Mol. Cell. Biol.** 19:182-193 (1999).

139. Kwon, Y. T. Lévy, F. and Varshavsky, A. Bivalent inhibitor of the N-end rule pathway. **J. Biol. Chem.** 274:18135-18139 (1999).

140. Xie, Y. and Varshavsky, A. The N-end rule pathway is required for import of histidine in yeast lacking the kinesin-like protein Cin8p. **Curr. Genet.** 36:113-123 (1999).

141. Xie, Y. and Varshavsky, A. The E2-E3 interactions in the N-end rule pathway: The RING-H2 finger of E3 is required for the synthesis of multiubiquitin chain. **EMBO J.** 18:6832-6844 (1999).

142. Suzuki, T. and Varshavsky, A. Degradation signals in the lysine-asparagine sequence space. **EMBO J.** 18:6017-6026 (1999).

143. Xie, Y. and Varshavsky, A. Physical association of ubiquitin ligases and the 26S proteasome. **Proc. Natl. Acad. Sci. USA** 97:2497-2502 (2000).

144. Kwon, Y. T., Balogh, S. A., Davydov, I. V., Kashina, A. S., Yoon, J. K., Xie, Y., Gaur, A., Hyde, L., Denenberg, V. H. and Varshavsky, A. Altered activity, social behavior, and spatial memory in mice lacking the NTAN1p amidase and the asparagine branch of the N-end rule pathway. **Mol. Cell. Biol.** 20:4135-4148 (2000).

145. Turner, G. C., Du, F. and Varshavsky, A. Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway. **Nature** 405:579-583 (2000).

146. Varshavsky, A. The ubiquitin system and the N-end rule pathway. **Biol. Chem.** 381:779-789 (2000).

149. Davydov, I. V. and Varshavsky, A. RGS4 is arginylated and degraded by the N-end rule pathway *in vitro*. **J. Biol. Chem.** 275:22931-22941 (2000).

150. Hershko, A., Ciechanover, A. and Varshavsky, A. The ubiquitin system. **Nature Medicine** 6:1073-1081 (2000). *(Three essays about early days of the ubiquitin field.)*

151. Xie, Y. and Varshavsky, A. RPN4 is a ligand, substrate, and transcriptional regulator of the 26S proteasome: a negative feedback circuit. **Proc. Natl. Acad. Sci. USA** 98, 3056-3061 (2001).

152. Rao, H., Uhlmann, F., Nasmyth, K. and Varshavsky, A. Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability. **Nature** 410:955-960 (2001).

153. Varshavsky, A. Recent studies of the ubiquitin system and the N-end rule pathway. **Harvey Lectures** (in press).

154. Varshavsky, A. Proteolysis. In **Encyclopedia of Genetics** (Academic Press, NY), in press.

155. Varshavsky, A. Ubiquitin. In **Encyclopedia of Genetics** (Academic Press, NY), in press.

U. S. Patents

1. Varshavsky, A. *Gene amplification assay for detecting tumor promoters.* Issued Apr. 10, 1984.
2. Snapka, R. M., Kwok, K. S., Bernard, J. A., Harling, O. R. and Varshavsky, A. *Indirect labeling method for post-separation detection of chemical compounds.* Issued Dec. 10, 1991.
3. Bachmair, A., Finley, D. and Varshavsky, A. *Methods for generating desired amino-terminal residues in proteins.* Issued Mar. 3, 1992.
4. Varshavsky, A., Johnson, E. S., Gonda, D. K. and Hochstrasser, M. *Methods for trans-destabilization of specific proteins in vivo and DNA molecules useful thereof.* Issued Jun. 16, 1992.
5. Bachmair, A., Finley, D. and Varshavsky, A. *Methods for producing proteins and polypeptides using ubiquitin fusions.* Issued Jul. 21, 1992.
6. Bachmair, A., Finley, D. and Varshavsky, A. *Methods for in vitro cleavage of ubiquitin fusion proteins.* Issued Mar. 23, 1993.
7. Baker, R. T., Tobias, J. W. and Varshavsky, A. *Nucleic acid encoding ubiquitin-specific proteases.* Issued May 18, 1993.
8. Tobias, J. W. and Varshavsky, A. *Ubiquitin-specific protease.* Issued Feb. 21, 1995.
9. Baker, R. T., Tobias, J. T. and Varshavsky, A. *Ubiquitin-specific proteases.* Issued Feb. 27, 1996.
10. Johnsson, N. and Varshavsky, A. *Split-ubiquitin protein sensor.* Issued Apr. 2, 1996.
11. Wu, P., Dohmen, R. J., Johnston, J. and Varshavsky, A. *Heat-inducible N-degron module.* Issued Jul. 23, 1996.
12. Johnston, J. and Varshavsky, A. *Inhibiting degradation of a degron-bearing protein.* Issued Jun. 9, 1998.
13. Baker, R. T., Gonda, D. K. and Varshavsky, A. *Inhibition of protein degradation in living cells with peptides.* Issued June 16, 1998.
14. Kwon, Y. T. and Varshavsky, A. *Nucleic acid encoding mammalian Ubr1.* Issued Jan. 19, 1999.

Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability

Hai Rao*, Frank Uhlmann†, Kim Nasmyth‡ & Alexander Varshavsky*

* *Division of Biology, California Institute of Technology, Pasadena, California 91125, USA*
† *Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK*
‡ *Research Institute of Molecular Pathology, Dr Bohr Gasse 7, A-1030 Vienna, Austria*

Cohesion between sister chromatids is established during DNA replication and depends on a protein complex called cohesin[1–7]. At the metaphase–anaphase transition in the yeast *Saccharomyces cerevisiae*, the *ESP1*-encoded protease separin cleaves SCC1, a subunit of cohesin with a relative molecular mass of 63,000 (M_r 63K)[8]. The resulting 33K carboxy-terminal fragment of SCC1 bears an amino-terminal arginine—a destabilizing residue in the N-end rule[9]. Here we show that the SCC1 fragment is short-lived ($t_{1/2} \approx 2$ min), being degraded by the ubiquitin/proteasome-dependent N-end rule pathway. Overexpression of a long-lived derivative of the SCC1 fragment is lethal. In *ubr1*Δ cells, which lack the N-end rule pathway[9], we found a highly increased frequency of chromosome loss. The bulk of increased chromosome loss in *ubr1*Δ cells is caused by metabolic stabilization of the ESP1-produced SCC1 fragment. This fragment is the first physiological substrate of the N-end rule pathway that is targeted through its N-terminal residue. A number of yeast proteins bear putative cleavage sites for the ESP1 separin, suggesting other physiological substrates and functions of the N-end rule pathway.

The sister chromatids of a replicated chromosome are pulled apart by the centromere-attached microtubules that emanate from spindle poles[1–6]. However, until the onset of anaphase the forces pulling sister chromatids apart are counteracted by cohesion that holds sisters together. A complex termed cohesin is essential for sister chromatid cohesion in both fungi and metazoans. In the yeast *S. cerevisiae*, cohesin contains four subunits, SMC1, SMC3, SCC1 (also called MCD1 and RAD21) and SCC3 (see refs 1–5). Perturbations of chromosome segregation are among the causes of human cancer and other diseases.

Previous work has shown that *S. cerevisiae* SCC1, a 63K subunit of cohesin, is cleaved at the metaphase–anaphase transition, in a reaction that requires the 187K ESP1, termed separin[8]. ESP1, a cysteine protease of a class that includes caspases[10], is activated at the onset of anaphase through ubiquitin (Ub)-dependent destruction of its inhibitory ligand PDS1 (see refs 1–3, 7). A protein substrate of the Ub system is conjugated to Ub through the action of enzymes E1, E2 and E3, with the degradation signal (degron) of the substrate recognized by E3. The resulting multiubiquitylated substrate is degraded by the 26S proteasome (see refs 11, 12).

The ESP1 separin cleaves the 566-residue SCC1 at two sites, 180 and 268 residues from its N terminus, the latter cleavage site being the principal one[8] (Fig. 1a). Mutations in both (but not single) SCC1 cleavage sites that inhibit the ESP1-mediated cleavage of SCC1 are lethal; they preclude sister chromatid separation[8,10]. The N-terminal Arg of ESP1-produced SCC1 fragments is a destabilizing residue in the N-end rule[9,13]. The degrons recognized by the N-end rule pathway, called N-degrons, consist of two determinants, a destabilizing N-terminal residue and an internal Lys residue[14]. UBR1, a 225K, RING-H2 finger-containing E3 of the N-end rule pathway[15], recognizes eight primary destabilizing N-terminal residues: Arg, Lys, His, Phe, Leu, Tyr, Trp and Ile. The other four destabilizing residues, Asp and Glu (secondary), and Asn and Gln (tertiary), function through their enzymatic modification and conjugation to Arg, a primary destabilizing residue[9]. UBR1 has at least three substrate-binding sites. The type 1 site is specific for N-terminal Arg, Lys and His. The type 2 site is specific for N-terminal Phe, Leu, Trp, Tyr and Ile. The third site of UBR1 is specific for proteins such as CUP9 and GPA1, which bear internal (non-N-terminal) degrons[16]. The functions of the N-end rule pathway include the control of peptide import in *S. cerevisiae*, by regulated degradation of CUP9—a transcriptional repressor of the *PTR2*-encoded peptide transporter[16]. Engineered N-end rule substrates can be produced through the Ub fusion technique, in which a Ub-reporter fusion is cleaved, co-translationally, at the Ub-reporter junction by deubiquitylating enzymes (DUBs), yielding a reporter bearing a predetermined N-terminal residue[13,17]. Neither type 1 nor type 2 physiological N-end rule substrates have yet been identified in *S. cerevisiae*.

To determine whether the ESP1-produced SCC1 fragments[8] were degraded by the N-end rule pathway, we took advantage of the fact that dipeptides bearing destabilizing N-terminal residues can specifically inhibit the yeast N-end rule pathway *in vivo*, through their import into the cell and interaction with the type 1 or type 2 sites of UBR1 (ref. 16). *S. cerevisiae* K6843, in which the full-length, C-terminally Myc$_{18}$-epitope-tagged SCC1 (termed SCC1$^{\mathrm{m}}$) was expressed from the P$_{GAL1\text{-}10}$ promoter[8], were arrested in a metaphase-like state with nocodazole, followed by the induction of SCC1$^{\mathrm{m}}$ with galactose. The dipeptides Ala-Arg, Arg-Ala, Trp-Ala or Leu-Ala were added 30 min later, and the incubation was continued for another 30 min, followed by extraction of proteins and anti-Myc immunoblotting.

In the absence of added dipeptides, we could detect only uncleaved SCC1$^{\mathrm{m}}$ (Fig. 1b, lane 2). However, in the presence of Arg-Ala, bearing a type 1 destabilizing N-terminal residue, an additional, smaller species was observed, whose size was consistent with the ESP1-produced major C-terminal fragment of SCC1$^{\mathrm{m}}$, termed SCC1$^{269-566\mathrm{m}}$ (Fig. 1b, lane 4). This band was absent from cells treated either with Ala-Arg, which bore a stabilizing N-terminal residue, or with Trp-Ala or Leu-Ala, both of which bore a type 2 destabilizing N-terminal residue (Fig. 1b, lanes 1–6). Crucially, SCC1$^{269-566\mathrm{m}}$ could also observed in the absence of dipeptides, provided that the cells lacked *UBR1* and therefore could not degrade N-end rule substrates (Fig. 1b, lanes 4, 7). No SCC1 fragment derived from cleavage solely at the minor site (residues 180–181) was observed (Fig. 1b). We conclude, in agreement with an earlier inference[8], that the ESP1-mediated minor-site cleavage of SCC1 is either negligible or rapidly followed by cleavage at the major site.

To determine the *in vivo* half-life of Arg-SCC1$^{269-566}$, it was expressed as part of a fusion of the form $^{\mathrm{f}}$DHFR–Ub–Arg-SCC1$^{269-566\mathrm{f}}$, where 'f' denotes the N-terminal and C-terminal Flag epitopes linked, respectively, to mouse dihydrofolate reductase (DHFR) and Arg-SCC1$^{269-566}$ (Fig. 1a). DUBs co-translationally cleave this fusion at the C terminus of the Ub moiety, yielding the test protein Arg-SCC1$^{269-566\mathrm{f}}$ and the long-lived $^{\mathrm{f}}$DHFR–Ub reference protein, which serves as an internal control[14,16,18]. In agreement with dipeptide inhibition experiments (Fig. 1b), Arg-SCC1$^{269-566\mathrm{f}}$ was short-lived in *UBR1* cells ($t_{1/2} \approx 2$ min) (Fig. 1c, lanes 1–4). By contrast, the same Arg-SCC1$^{269-566\mathrm{f}}$ protein was long-lived ($t_{1/2} > 2$ h) in *ubr1*Δ cells, indicating that its degradation in wild-type cells was mediated by the N-end rule pathway (Fig. 1c, lanes 5–8; and 1e). Arg-SCC1$^{269-566\mathrm{f}}$ was destroyed particularly rapidly either during or shortly after its synthesis, that is, during the pulse. Specifically, nearly 90% of Arg-SCC1$^{269-566\mathrm{f}}$ was degraded by the N-end rule pathway during the 3-min pulse; the rest of pulse-labelled Arg-SCC1$^{269-566\mathrm{f}}$ was destroyed more slowly ($t_{1/2} \approx 2$ min) (Fig. 1e). If the N-terminal Arg of Arg-SCC1$^{269-566\mathrm{f}}$ was converted to Met, a stabilizing residue in the N-end rule, the resulting Met-

SCC1$^{269-566f}$ (Fig. 1a) was long-lived in both *UBR1* and *ubr1*Δ cells, in contrast to Arg-SCC1$^{269-566f}$ (Fig. 1c, d). We conclude that Arg-SCC1$^{269-566f}$ is targeted for degradation through its Arg-bearing N-degron, rather than through an internal (non-N-terminal) degradation signal.

Overexpression of the N-terminal SCC1 fragment (residues 1–268; Fig. 1a) in either *UBR1* or *ubr1*Δ cells did not significantly alter cell growth (data not shown). In contrast, the analogous expression of Arg-SCC1$^{269-566f}$ was lethal in *ubr1*Δ cells (Fig. 2a). In *UBR1* cells, where Arg-SCC1$^{269-566f}$ was short-lived (Fig. 2a), its overexpression was not lethal but slowed growth (data not shown). Overexpression of the Met-SCC1$^{269-566f}$ was lethal in both *UBR1* and *ubr1*Δ cells, consistent with metabolic stability of this SCC1 derivative in both genetic backgrounds (Fig. 2a).

To examine toxicity of the long-lived Met-SCC1$^{269-566f}$ in more detail, we transiently overexpressed Met-SCC1$^{269-566f}$ from the P_{GAL1} promoter in cells that were arrested either before the S phase (with α-factor), or in the S phase (with hydroxyurea), or in a metaphase-like state (with nocodazole). Transient overexpression of Met-SCC1$^{269-566f}$ was largely nontoxic to cells in G1 and S phases but highly toxic to mitotic (nocodazole-arrested) cells (Fig. 2b), indicating that elevated levels of SCC1$^{269-566}$ may interfere with a process that takes place specifically at metaphase–anaphase. Control immunoblots confirmed that the levels of Met-SCC1$^{269-566f}$ were about equal in cells arrested at either of the three phases of the cell cycle (data not shown). Furthermore, we found that cells remained stably arrested in the presence of nocodazole even after overexpression of the SCC1 fragment (data not shown), thus making it unlikely that abnormally high levels of this fragment abrogated the MAD2-dependent checkpoint and the cell-cycle arrest induced by nocodazole. It is also possible that selective toxicity of the SCC1 fragment in nocodazole-arrested cells is caused by this fragment acting as a competitive inhibitor of the ESP1 protease.

Full-length SCC1 interacts with SMC1, a DNA-binding subunit of cohesin[1,2]. Therefore one cause of lethality of overexpressed SCC1$^{269-566}$ may be a residual (and functionally perturbing) affinity of this fragment for the rest of the cohesin complex. In this model,

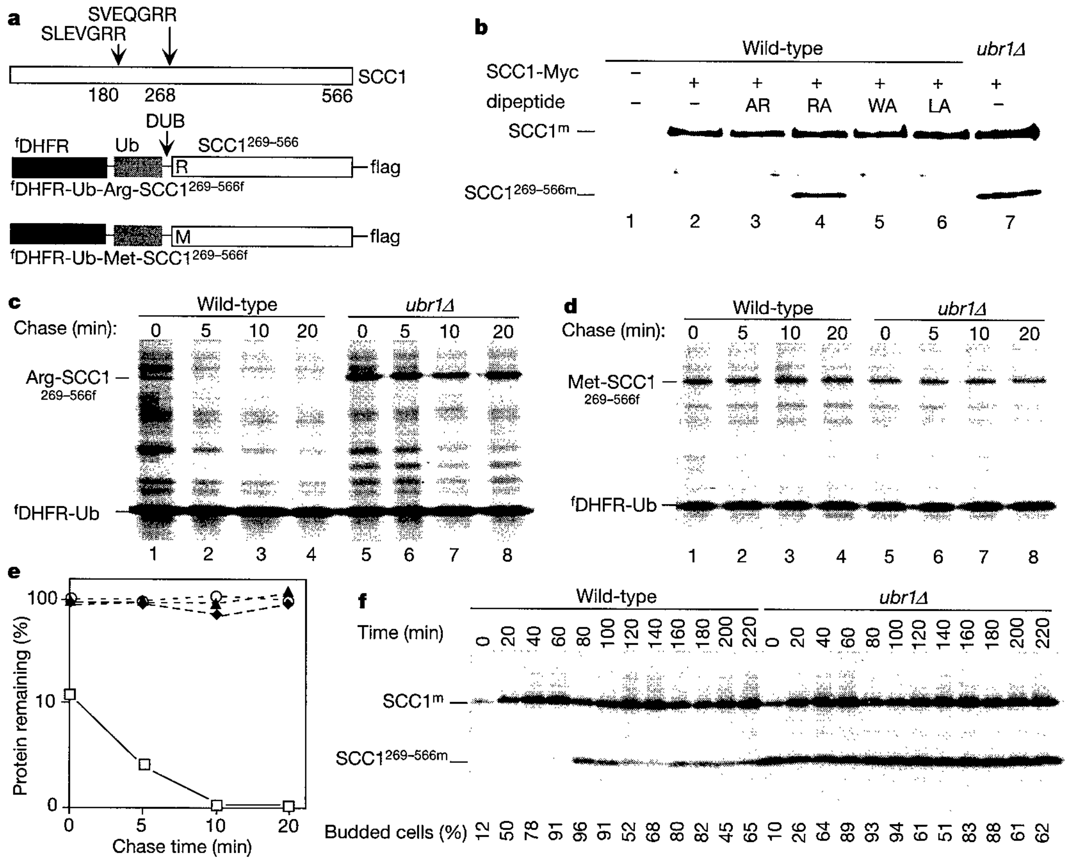


Figure 1 The SCC1 fragment is a substrate of the N-end rule pathway. **a**, The *in vivo* cleavage sites in *S. cerevisiae* SCC1 (ref. 8), and the UPR-type fusions that produce, on co-translational cleavage by DUBs, the Flag-tagged reference protein ᶠDHFR–Ub and the Flag-tagged test proteins Arg-SCC1$^{269-566f}$ or Met-SCC1$^{269-566f}$. **b**, Detection of SCC1 fragment. Control *S. cerevisiae* cells and cells expressing SCC1^{m} were incubated with specific dipeptides, followed by immunoblotting with anti-Myc antibody. Lane 1, control cells (lacking SCC1^{m}). Lane 2, cells expressing SCC1^{m} in the absence of added dipeptide. Lanes 3–6, as lane 2 but in the presence of Ala-Arg (AR), Arg-Ala (RA), Trp-Ala (WA) and Leu-Ala (LA), respectively. Lane 7, *ubr1*Δ cells expressing SCC1^{m} in the absence of dipeptides. Arrows indicate bands of SCC1^{m} and SCC1$^{269-566m}$. **c**, Pulse–chase analysis of Arg-SCC1$^{269-566f}$. Cells expressing ᶠDHFR–Ub–Arg-SCC1$^{269-566f}$ were labelled for 3 min with [^{35}S]methionine, followed by a chase for 5, 10 or 20 min, and SDS–PAGE analysis of anti-Flag immunoprecipitates. Lanes 1–4, pulse chase in *UBR1* cells. Lanes 5–8, pulse chase in *ubr1*Δ cells. Arrows indicate bands of Arg-SCC1$^{269-566f}$ and the reference protein ᶠDHFR–Ub. **d**, As in **c**, but with cells expressing Met-SCC1$^{269-566f}$, derived from ᶠDHFR–Ub–Met-SCC1$^{269-566f}$. **e**, Quantification of results in **c** and **d** by PhosphorImager. Squares, Arg-SCC1$^{269-566f}$ in JD52 (*UBR1*) cells; diamonds, Arg-SCC1$^{269-566f}$ in JD55 (*ubr1*Δ) cells; circles, Met-SCC1$^{269-566f}$ in JD52 (*UBR1*) cells; triangles, Met-SCC1$^{269-566f}$ in JD55 (*ubr1*Δ) cells. The amounts of ^{35}S in the test proteins, relative to ^{35}S in the ᶠDHFR–Ub reference protein, were plotted as percentages of this ratio for Met-SCC1$^{269-566f}$ in *UBR1* cells at time zero. **f**, Immunoblot analysis, using anti-Myc antibody, of Arg-SCC1$^{269-566m}$ produced through ESP1-mediated cleavage of normally expressed, C-terminally Myc-tagged SCC1^{m} in wild-type (*UBR1*) and *ubr1*Δ cultures synchronized with α-factor, as described[8]. The strains used were *S. cerevisiae* Y92 (*MAT***a** *Scc1Myc18*)[8] and Y238 (*MAT***a** *Scc1Myc18 ubr1*Δ). Shown are the time after removal of α-factor (top) and the content of budded (post-G1 phase) cells (bottom) in *UBR1* and congenic *ubr1*Δ cultures.

the function of metabolic instability of the SCC1$^{269-566}$ fragment is to ensure its removal from the cohesin complex, allowing rebinding of intact SCC1 in the next cell cycle. To determine whether Met-SCC1$^{269-566f}$ could, in fact, interact with SMC1, extracts from cells expressing Flag-tagged Met-SCC1$^{269-566f}$, either alone or together with RGS/His$_6$-tagged SMC1, were precipitated with anti-Flag antibody, followed by SDS polyacrylamide gel electrophoresis (SDS–PAGE) of the immunoprecipitates and immunoblotting with either anti-Flag or anti-RGS/His$_6$ antibodies. We found that Met-SCC1$^{269-566f}$ was specifically co-immunoprecipitated with the SMC1 (Fig. 2c). However, overexpression of SMC1, from the P$_{GAL1}$ promoter and low copy plasmid, did not suppress the lethal effect of Met-SCC1$^{269-566f}$ overexpression in *UBR1* cells (data not shown). Thus, the toxicity of overexpressed SCC1$^{269-566}$ may not be caused exclusively by sequestration of cohesin complexes.

Given the above results (Figs 1 and 2), might *ubr1Δ* cells, which lack the N-end rule pathway and in which the Arg-SCC1 fragment is long-lived, be impaired in the maintenance of chromosome stability? We used a colony sectoring assay[19] to compare the rates of chromosome loss in *ubr1Δ* and *UBR1* cells (Fig. 3a). Remarkably, whereas less than 0.5% of *UBR1* colonies contained red sectors, ~22% of the *ubr1Δ* colonies contained such sectors (Fig. 3a; and data not shown), revealing a previously unknown phenotype of chromosome instability in cells lacking the N-end rule pathway. The frequency of distinct red sectors (Fig. 3a) is roughly proportional to the frequency of cells that suffered the loss of a (non-essential) marked chromosome during the growth of colonies. We conclude that *ubr1Δ* cells lose chromosomes about 100 times more frequently than their congenic *UBR1* counterparts.

To determine how much of the chromosome-loss phenotype of *ubr1Δ* cells was caused by their inability to degrade specifically the Arg-SCC1 fragment (as distinguished from other N-end rule substrates), we constructed a *S. cerevisiae* strain (in the background of a sectoring strain[19]) in which wild-type *SCC1* was replaced by an allele (expressed from the native P$_{SCC1}$ promoter) that encoded SCC1 bearing at its C terminus three tandem haemagglutinin A (HA) tags and lacking the minor ESP1 cleavage site (Fig. 1a), owing to an Arg to Glu mutation at position 180. In this strain, the only ESP1-produced SCC1 fragment was Arg-SCC1$^{269-566h}$. We also constructed a derivative of this strain in which Arg-269 of SCC1–HA was replaced with Gly. The resulting ESP1-produced fragment of SCC1 was the long-lived Gly-SCC1$^{269-566h}$ bearing a stabilizing N-terminal residue.

We found that the efficiency of cleavage at the thus altered cleavage site of SCC1 was indistinguishable from that at the wild-type cleavage site (Fig. 3b). In the strain producing the wild-type (short-lived) Arg-SCC1$^{269-566}$ fragment, the frequency of chromosome loss was indistinguishable from that in the congenic wild-type strain (Fig. 3a; and data not shown). By contrast, in the strain in which the ESP1-produced Gly-SCC1$^{269-566h}$ fragment bore a stabilizing N-terminal residue, a greatly increased frequency of chromosome loss was observed, similarly to congenic *ubr1Δ* cells expressing wild-type SCC1 (Fig. 3a; and data not shown). We conclude that the bulk of increased chromosome loss in *ubr1Δ* cells is caused by metabolic stabilization of the ESP1-produced SCC1 fragment.

Although the steady-state level of ESP1-generated Arg-SCC1$^{269-566h}$ was significantly (~3-fold) lower in wild-type (*UBR1*) than in *ubr1Δ* cells (Fig. 3b, lanes 2 and 3), the observed difference was less than would be expected from the extremely short (~2 min) half-life of overproduced Arg-SCC1$^{269-566}$ in *UBR1* cells (Fig. 1c, d). To assess the influence of the N-end rule pathway on the concentration of endogenous, ESP1-produced Arg-SCC1$^{269-566}$ fragment, we monitored its levels by immunoblotting as a function of time,

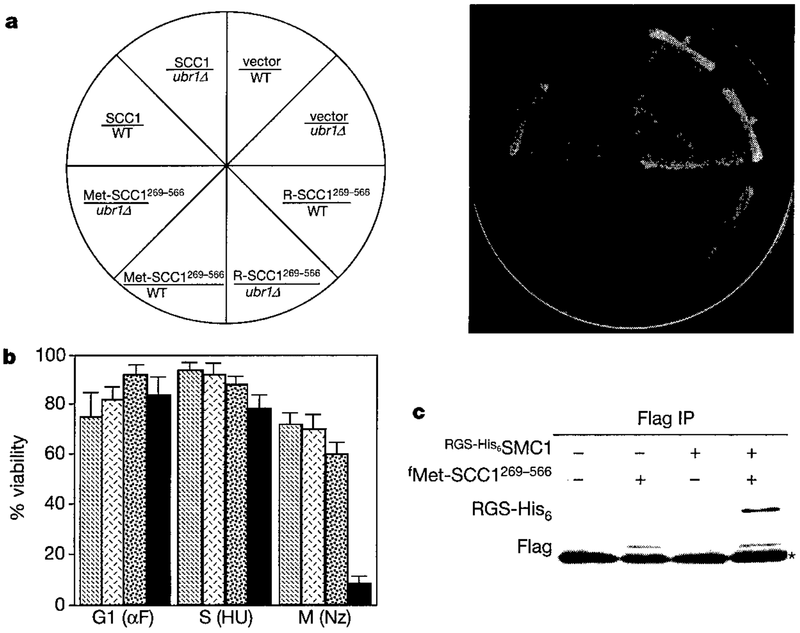
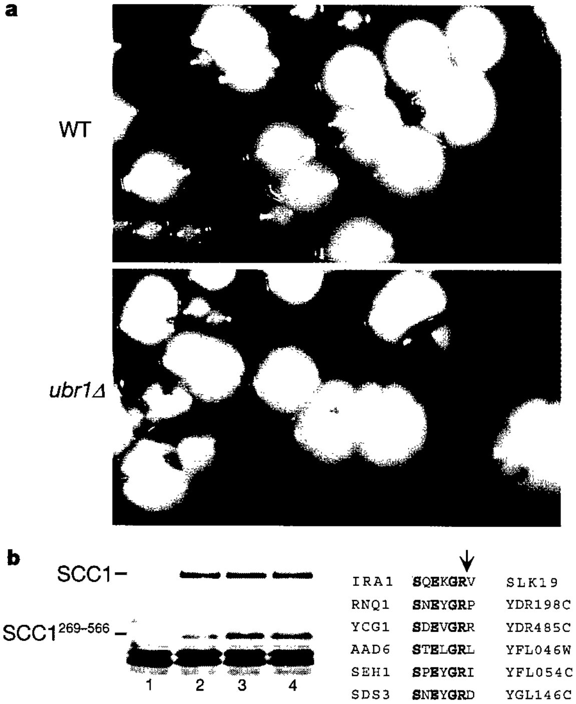


Figure 2 Effects of overexpressing SCC1 and its fragments in *UBR1* and *ubr1Δ* cells. **a**, Overexpression of long-lived SCC1$^{269-566}$ is lethal. Strains K6843 (*UBR1*), JD52 (*UBR1*) or JD55 (*ubr1Δ*) carrying plasmids that overexpressed, from the P$_{GAL1}$ promoter, full-length SCC1, Arg-SCC1$^{269-566}$ or Met-SCC1$^{269-566f}$ (left) were streaked either on dextrose-containing (SD) plates (not shown), on which all strains grew, or on galactose-containing (SG) plates (right). The plates were incubated for 3 d at 30 °C. **b**, Plating efficiency of *UBR1* cells that transiently expressed Met-SCC1$^{269-566}$ while being transiently arrested in late G1 phase (with α-factor, αF), in S phase (with hydroxyurea, HU), or in a metaphase-like state (with nocodazole, Nz). For each set of assays: the first bar refers to wild-type (*UBR1*) AVY110 cells incubated in raffinose-containing medium (YP-R); the second to the same strain in galactose-containing YP medium (YP-G); the third to congenic AVY111 cells in YP-R that carried the ORF encoding Met-SCC1$^{269-566f}$ downstream from the P$_{GAL1}$ promoter; and the fourth to the same strain in YP-G. **c**, The SCC1$^{269-566}$ fragment interacts with SMC1. Extracts of *S. cerevisiae* expressing either Met-SCC1$^{269-566f}$ (Flag-tagged), or RGS/His$_6$-tagged SMC1, or both were precipitated with anti-Flag antibody, followed by SDS–PAGE and immunoblotting with either anti-Flag or anti-RGS/His$_6$. Asterisk denotes a protein crossreacting with anti-Flag antibody.

using synchronized cultures of *UBR1* and *ubr1Δ* cells. The levels of full-length SCC1 were regulated similarly in *UBR1* and *ubr1Δ* cells (Fig. 1f). In a *UBR1* culture, cells in G1 phase contained virtually no Arg-SCC1$^{269-566}$ fragment, which appeared transiently as cells entered anaphase (Fig. 1f). In striking contrast, *ubr1Δ* cells contained the Arg-SCC1$^{269-566}$ fragment throughout the cell cycle, and in addition at levels that significantly exceeded the peak level of Arg-SCC1$^{269-566}$ in *UBR1* cells (Fig. 1f).

Together, the short-half life of the Arg-SCC1$^{269-566}$ fragment (Fig. 1e), the toxicity of its long-lived counterpart (Fig. 2), the altered accumulation pattern and steady-state levels of Arg-SCC1$^{269-566}$ in *ubr1Δ* cells (Fig. 1f), the greatly increased chromosome loss in *ubr1Δ* cells (Fig. 3a), and the evidence that most of this loss is caused by metabolic stabilization of Arg-SCC1$^{269-566}$ allow the following conclusions. First, if the Arg-SCC1$^{269-566}$ fragment becomes long-lived (as it does in *ubr1Δ* cells), even the relatively low, wild-type level of its expression is detrimental to the fidelity of chromosome maintenance. Second, the metabolic stabilization of N-end rule substrates other than Arg-SCC1$^{269-566}$ seems not to contribute significantly to the elevated chromosome loss in *ubr1Δ* cells. As to the cause of toxicity of overexpressed SCC1$^{269-566}$ fragment, one possibility, suggested by high SCC1$^{269-566}$ levels in G1-phase *ubr1Δ* cells (Fig. 1f), is that the fragment interferes with re-formation of cohesin complexes that are required for the establishment of cohesion during the S phase.



c

				IRA1	SQEKGRV	SLK19	SNEEGRN
				RNQ1	SNEYGRP	YDR198C	SPETGRG
				YCG1	SDEVGRR	YDR485C	SKEEGRS
				AAD6	STELGRL	YFL046W	SLEKGRI
				SEH1	SPEYGRI	YFL054C	SYESGRS
				SDS3	SNEYGRD	YGL146C	SHERGRQ
				CCP1	SVEKGRS	YHR080C	SEENGRQ
SCC1	SLEVGRR	YBR108W	STENGRL	SRL2	SAEDGRL	YJR088C	SRELGRN
SCC1	SVEQGRR	YBR147W	SLELGRT	YSH1	SNEVGRS	YKL146W	SREHGRH
REC8	SVERGRK	YBR292C	SQEYGRI	HMG1	SEEGGRL	YMR144W	SPEMGRT
REC8	SHEYGRK	YDL117W	STENGRQ	SPO20	SSEYGRY	YPL138C	SEEHGRE
CPR8	SEEAGRL	YDL211C	SFENGRD	EMP47	SLEEGRI		

Figure 3 Increased chromosome instability in *ubr1Δ S. cerevisiae*. **a**, Sectoring assay for the chromosome loss using the *SUP11*-marked YPH277 (*UBR1*) strain and its *ubr1Δ* derivative (see Methods). **b**, Immunoblot analysis of Arg-SCC1$^{269-566h}$ and Gly-SCC1$^{269-566h}$, produced through ESP1-mediated cleavage of normally expressed, C-terminally HA-tagged SCC1 in wild-type and *ubr1Δ* strains. Yeast cultures ($A_{600} \approx 0.4$) were processed for immunoblotting with anti-HA antibody (Covance, Berkeley, CA). Lane 1, YPH277 (control) cells lacking the HA-tagged SCC1. Lane 2, AVY129 (*UBR1*) cells expressing normal levels of SCC1^{R180E}–HA. Lane 3, as lane 2 but with AVY132 (*ubr1Δ*) cells. Lane 4, as lane 2, but with SCC1$^{R180E,\ R269G}$–HA, bearing the Arg to Gly mutation at position 269. The bands of full-length SCC1–HA and its ESP1-produced SCC1$^{269-566h}$ fragments are indicated on the left. **c**, *S. cerevisiae* proteins containing putative cleavage sites for the ESP1 separin. The query SxExGRx and the Pattern Match program (http://genome-www2.stanford.edu/cgi-bin/SGD/PATMATCH/nph-patmatch) were used to search the *S. cerevisiae* genome database. Arrows indicate the inferred sites of cleavage by ESP1.

The increased probability of chromosome loss in *ubr1Δ* cells, although readily detectable through the sectoring assay (Fig. 3a), is still below levels that would cause massive cell death during growth of a culture. In particular, no significant differences were observed between congenic *ubr1Δ* and *UBR1* cells in the kinetics of cell-cycle progression (data not shown). However, a further significant increase in the level of long-lived SCC1$^{269-566}$ (achieved through its transcriptional overexpression) was found to be lethal (Fig. 2a). A parsimonious interpretation is that the latter effect is mechanistically similar to the much less severe, viability-compatible chromosome-loss phenotype of *ubr1Δ* cells, in which SCC1$^{269-566}$ becomes long-lived but is not overexpressed otherwise. In this interpretation, the frequency of chromosome loss in cells that overexpress long-lived SCC1$^{269-566}$ becomes high enough for them to lose a chromosome nearly every division. Light-microscopic observations indicated that individual cells divided once or twice after overexpression of Met-SCC1$^{269-566}$, followed by growth arrest and a varying terminal phenotype (data not shown).

Some physiological N-end rule substrates other than SCC1 may also be produced by the ESP1 separin. The consensus sequence of the two ESP1 cleavage sites in SCC1 is SxExGR↓R (Fig. 1a). *S. cerevisiae* REC8, the meiosis-specific counterpart of SCC1, contains two near-matches to this consensus[8]. REC8 is cleaved by ESP1 during yeast meiosis, similarly to SCC1 in vegetative cells[20], and potential cleavage sites are also present in *Schizosaccharomyces pombe* RAD21 and REC8, the fission yeast homologues of SCC1 and REC8 (ref. 8). Simultaneous (but not single) alterations of these sites in *S. pombe* RAD21 block chromosome segregation, similarly to the results in *S. cerevisiae*[21]. A shorter consensus, SxExGR↓, derived from comparisons of the ESP1 cleavage sites in the budding and fission yeast SCC1 homologues[8], was used to search the *S. cerevisiae* genome database, yielding 29 putative substrates of ESP1 other than SCC1 and REC8 (Fig. 3c). The N-terminal residues of the predicted C-terminal fragments of these proteins encompass nearly the entire set of 12 residues that are destabilizing in the N-end rule, including tertiary and secondary destabilizing residues (Fig. 3c).

Until now, the known substrates of the N-end rule pathway that are recognized through an N-degron were either polyprotein-derived viral proteins[9,22] or proteins secreted by an intracytoplasmic bacterium such as *Listeria monocytogenes* into the mammalian cell's cytosol[23]. Processing proteases that convert pro-N-degrons into N-degrons have long been envisioned as regulated entry points to the N-end rule pathway[9,13]. The identification of ESP1 as one such entry point, and of SCC1 as the first N-degron-based physiological N-end rule substrate in yeast, marks the initial understanding of connections between the N-end rule and the ESP1 separin. The discovery of increased chromosome instability in *ubr1Δ* cells suggests that chromosome aneuploidy and related mutator phenotypes characteristic of mammalian cancer cells may also involve perturbations of the N-end rule pathway. Because ESP1 homologues are present in most if not all eukaryotes[24], identifying ESP1 substrates in these organisms should be an effective way to pinpoint physiological N-end rule substrates, as well as additional, currently unknown, functions of the N-end rule pathway. ☐

Methods

Yeast strains and plasmids

Strains *S. cerevisiae* JD52 (*MATa lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-3,112*) and JD55 (*ubr1::HIS3* in the JD52 background) have been described[25]. *S. cerevisiae* K6843 (*MATa bar1 ade2-1 trp1-1 can1-100 leu2-3,112 his3-11,15 Gal-Scc1Myc18::URA3*) expressed full-length, C-terminally Myc$_{18}$-tagged SCC1 from the $P_{GAL1-10}$ promoter[8]. We produced AVY115 (*ubr1::LEU2*), a UBR1-lacking derivative of K6843, using standard techniques. The plasmids p426CUPR-SCC1, p426CUPR-SCC1R, p426CUPR-SCC1M and p426CUPR-SCC1N expressed, respectively, full-length SCC1^{f}, Arg-SCC1$^{269-566f}$, Met-SCC1$^{269-566f}$ and SCC1^{1-268f} from the P_{CUP1} promoter as UPR[14] fusions of the form ^{f}DHFR–Ub–(SCC1 derivative^{f}) (Fig. 1a). The Ub moiety of these fusions bore a Lys to Arg alteration at position 48.

^{f}DHFR–Ub–Arg-SCC1$^{269-566f}$ and ^{f}DHFR–Ub–Met-SCC1$^{269-566f}$ were expressed from the P_{GAL1} promoter in the plasmids p416UPRSCC1R and p416UPRSCC1M, respectively, which were derived from the low-copy vector p416Gal1 (ref. 26). The N-terminally RGS/His$_6$-tagged SMC1 was expressed from the P_{GAL1} promoter in plasmid p415GalSMC1, derived from p415Gal1 (ref. 26).

We cloned a fragment encoding Met-SCC1$^{269-566f}$ downstream from the P_{GAL1} promoter into pRS305 (ref. 25). The resulting plasmid was cut with *Xcm*I within LEU2 and transformed into *S. cerevisiae* AVY110 (*MAT***a** *bar1-1 ade2-1 leu2-3,112 ura3-1*), yielding AVY111 (*leu2::*P_{GAL1}*-Met-SCC1*$^{269-566f}$*::LEU2*). UBR1 was deleted from the sectoring-assay strain YPH277 (*MAT***a** *ura3-52 ade2-101 trp1-Δ1 leu2-Δ1 CFVII* (*RAD2.d*) *URA3 SUP11*)[19], yielding YPH277HR1 (*ubr1::LEU2*). We also constructed, in the background of the pYIplac128 vector, the alleles of SCC1 encoding the triple-HA tag at the SCC1 C terminus[10] and either the Arg to Glu mutation at position 180 or, in addition, the Arg to Gly mutation at position 269. The resulting plasmids were cut with *Eco*RV within *LEU2* and transformed into the diploid sectoring-assay strain AVY125 (*MAT***a***/MAT*α *SCC1/scc1::Kan*^{R} *ura3-52/ura3-52 ade2-101/ade2-101 TRP1/trp1-Δ1 leu2-Δ1/ leu2-Δ1 CFVII* (*RAD2.d*) *URA3 SUP11*)[19] lacking one of two copies of SCC1. Haploid derivatives of the resulting transformants, verified for the presence of the intended SCC1 alleles, were the strains AVY129 (*MAT***a** *scc1::Kan*^{R} *leu2::SCC1*^{R180E}*::LEU2 ura3-52 ade2-101 trp1-Δ1 CFVII* (*RAD2.d*) *URA3 SUP11*) and AVY130 (*MAT***a** *scc1::Kan*^{R} *leu2::SCC1*$^{R180E\ R269G}$*::LEU2 ura3-52 ade2-101 trp1-Δ1 CFVII* (*RAD2.d*) *URA3 SUP11*).

Experiments with dipeptides

We grew *S. cerevisiae* K6843 (*UBR1*)[8] and AVY115 (*ubr1*Δ) at 30 °C in YP (yeast-peptone) medium containing 2% raffinose as the carbon source to an absorbance at 600 nm (A_{600}) ≈ 0.5. Cells were arrested at metaphase by nocodazole (Sigma), at 15 μg ml^{-1}. The $P_{GAL1-10}$ promoter was induced by the addition of galactose to 2% for 30 min. We added specific dipeptides (Sigma) to 5 mM, and incubated 2-ml cultures for another 30 min, followed by the addition of an equal volume of 0.3 M NaCl, 0.1 M NaF, 20 mM EDTA, 2 mM NaN_3 (pH 8.2). Cells were pelleted by centrifugation, resuspended in 0.2 ml of the lysis buffer (10% glycerol, 50 mM NaOH, 2% SDS, 5% β-mercaptoethanol), and heated at 100 °C for 5 min. The suspension was titrated with 1 M HCl to pH 7, and centrifuged at 12,000*g* for 2 min. Samples of the supernatants (10 μl) were subjected to SDS–PAGE (10% gel), followed by immunoblotting with a 1:1000 dilution of anti-Myc 9E10 antibody (Berkeley Antibody Co.) The blot was incubated with a 1:2000 dilution of the goat anti-mouse horseradish peroxidase (HRP) conjugate, and was developed using ECL reagents (Amersham).

Pulse-chase analysis

S. cerevisiae JD52 and JD55, carrying plasmids that expressed ^{f}DHFR–Ub–Arg-SCC1$^{269-566f}$ or ^{f}DHFR–Ub–Met-SCC1$^{269-566f}$ from the P_{GAL1} promoter, were grown at 30 °C to A_{600} ≈ 1 in SR(−ura) medium with auxotrophic supplements and 2% raffinose as the carbon source. P_{GAL1} was induced by adding galactose, to 2% (SRG(−ura) medium), for 1 h. Cells from a 20-ml culture were gathered by centrifugation, washed with 0.8 ml SRG(−ura), resuspended in 0.4 ml of SRG(−ura), and labelled for 3 min with 0.16 mCi of ^{35}S-Express (NEN), followed by centrifugation and resuspension of cells in SD(−ura) containing 4 mM L-methionine and 2 mM L-cysteine. We took 0.1 ml samples at the indicated time points and processed them for immunoprecipitation with anti-Flag M2 agarose beads (Sigma), followed by SDS–PAGE, as described[25].

Experiments with growth inhibitors

We grew *S. cerevisiae* AVY111, which expresses Met-SCC1$^{269-566f}$ from the P_{GAL1} promoter, and the congenic parental strain AVY110 at 30 °C in YP medium containing 2% raffinose as the carbon source to A_{600} ≈ 0.5. To separate samples of the two cultures, were added α-factor (to 1 μg ml^{-1}), hydroxyurea (to 0.1 M) or nocodazole (to 15 μg ml^{-1}), followed by incubation for 2.5 h. One half of each of the resulting samples was then incubated further in the same medium for 1.5 h; the other half had galactose added to 2%, followed by a 1.5-h incubation. Cells were gathered by centrifugation, washed with water, and plated onto dextrose-containing synthetic-media (SD) plates.

Co-immunoprecipitation/immunoblotting assay

JD52 cells carrying two vectors (pRS415 and pRS416[25]), or pRS415 and p416UPRSCC1M (expressing Flag-tagged Met-SCC1$^{269-566f}$), or pRS416 and p415GalSMC1 (expressing RGS/His$_6$-tagged SCM1), or p415GalSMC1 and p416UPRSCC1M together, were grown in the raffinose-containing SR medium to A_{600} ≈ 1. Galactose was added to 2%, followed by incubation for 2 h, preparation of extracts[25], immunoprecipitation with anti-Flag M2 beads, SDS–PAGE (10% gel), and immunoblotting, separately, with anti-Flag and anti-RGS/His$_6$ antibodies (Qiagen).

Sectoring assay

S. cerevisiae YPH277 contained the defective *ade2-101* ochre allele of *ADE2*, which was suppressed by the SUP11 suppressor transfer RNA gene, carried on a non-essential, centromere-containing, ~90-kilobase chromosome fragment[19]. YPH277 cells that retained the *SUP11*-containing chromosome formed white colonies; the loss of this chromosome resulted in a red sector or sectors. The *UBR1* gene was deleted in the YPH277 background, and the resulting *ubr1*Δ strain YPH277HR1 was compared with YPH277. Cells were grown in the uracil-lacking SD medium that retained the *SUP11*-containing chromosome, followed by plating on YPD plates.

Received 20 November 2000; accepted 17 January 2001.

1. Nasmyth, K., Peters, J. M. & Uhlmann, F. Splitting the chromosome: cutting the ties that bind sister chromatids. *Science* **288,** 1379–1384 (2000).
2. Koshland, D. & Guacci, V. Sister chromatid cohesion: the beginning of a long and beautiful relationship. *Curr. Opin. Cell Biol.* **12,** 297–301 (2000).
3. Yanagida, M. Cell cycle mechanisms of sister chromatid separation: roles of Cut1/separin and Cut2/securin. *Genes Cells* **5,** 1–8 (2000).
4. Hirano, T. Chromosome cohesion, condensation, and separation. *Annu. Rev. Biochem.* **69,** 115–144 (2000).
5. Dej, K. J. & Orr-Weaver, T. L. Separation anxiety at the centromere. *Trends Cell Biol.* **10,** 392–399 (2000).
6. Pidoux, A. L. & Allshire, R. C. Centromeres: getting a grip of chromosomes. *Curr. Opin. Cell Biol.* **12,** 308–319 (2000).
7. Biggins, S. & Murray, A. W. Sister chromatid cohesion in mitosis. *Curr. Opin. Genet. Dev.* **9,** 230–236 (1999).
8. Uhlmann, F., Lottspeich, F. & Nasmyth, K. Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. *Nature* **400,** 37–42 (1999).
9. Varshavsky, A. The N-end rule: functions, mysteries, uses. *Proc. Natl Acad. Sci. USA* **93,** 12142–12149 (1996).
10. Uhlmann, F., Wernic, D., Poupart, M.-A., Koonin, E. V. & Nasmyth, K. Cleavage of cohesin by the CD-clan protease separin triggers anaphase in yeast. *Cell* **103,** 375–386 (2000).
11. Varshavsky, A. The ubiquitin system. *Trends Biochem. Sci.* **22,** 383–387 (1997).
12. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **76,** 425–479 (1998).
13. Bachmair, A., Finley, D. & Varshavsky, A. *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* **234,** 179–186 (1986).
14. Suzuki, T. & Varshavsky, A. Degradation signals in the lysine-asparagine sequence space. *EMBO J.* **18,** 6017–6026 (1999).
15. Xie, Y. & Varshavsky, A. The E2-E3 interaction in the N-end rule pathway: the RING-H2 finger of E3 is required for the synthesis of multiubiquitin chain. *EMBO J.* **18,** 6832–6844 (1999).
16. Turner, G. C., Du, F. & Varshavsky, A. Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway. *Nature* **405,** 579–583 (2000).
17. Varshavsky, A. Ubiquitin fusion technique and its descendants. *Methods Enzymol.* **327,** 578–593 (2000).
18. Turner, G. C. & Varshavsky, A. Detecting and measuring cotranslational protein degradation *in vivo. Science* **289,** 2117–2120 (2000).
19. Spencer, F., Gerring, S. L., Connelly, C. & Hieter, P. Mitotic chromosome transmission fidelity mutants in *Saccharomyces cerevisiae. Genetics* **124,** 237–249 (1990).
20. Buonomo, S. B. C. *et al.* Disjunction of homologous chromosomes in meiosis I depends on proteolytic cleavage of the meiotic cohesin Rec8 by separin. *Cell* **103,** 387–398 (2000).
21. Tomonaga, T. *et al.* Characterization of fission yeast cohesin: essential anaphase proteolysis of Rad21 phosphorylated in the S phase. *Genes Dev.* **14,** 2757–2770 (2000).
22. Mulder, L. C. F. & Muesing, M. A. Degradation of HIV-1 integrase by the N-end rule pathway. *J. Biol. Chem.* **275,** 29749–29753 (2000).
23. Sijts, A. J., Pilip, I. & Pamer, E. G. The *Listeria monocytogenes*-secreted p60 protein is an N-end rule substrate in the cytosol of infected cells. Implications for major histocompatibility complex class I antigen processing of bacterial proteins. *J. Biol. Chem.* **272,** 19261–19268 (1997).
24. Waizenegger, I. C., Hauf, S., Meinke, A. & Peters, J.-M. Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. *Cell* **103,** 399–410 (2000).
25. Ghislain, M., Dohmen, R. J., Levy, F. & Varshavsky, A. Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in *Saccharomyces cerevisiae. EMBO. J.* **15,** 4884–4899 (1996).
26. Mumberg, D., Muller, R. & Funk, M. Regulatable promoters of *Saccharomyces cerevisiae*—comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.* **22,** 5767–5768 (1994).

Acknowledgements

We are grateful to G. Turner, H.-R. Wang, F. Du, V. Ellison, A. Murray, D. Morgan, B. Stillman and P. Hieter for strains and plasmids. We thank M. Budd, W. Shuo and members of the Varshavsky laboratory, particularly F. Du and G. Turner, for helpful discussions. This work was supported by grants to A.V. from the NIH. H.R. is a Fellow of the Leukemia and Lymphoma Society. K.N. was supported by the Austrian Industrial Research Promotion Fund.

Correspondence and requests for materials should be addressed to A.V. (e-mail: avarsh@caltech.edu).

RPN4 is a ligand, substrate, and transcriptional regulator of the 26S proteasome: A negative feedback circuit

Youming Xie and Alexander Varshavsky*

Division of Biology, California Institute of Technology, Pasadena, CA 91125

Contributed by Alexander Varshavsky, January 16, 2001

The RPN4 (SON1, UFD5) protein of the yeast *Saccharomyces cerevisiae* is required for normal levels of intracellular proteolysis. RPN4 is a transcriptional activator of genes encoding proteasomal subunits. Here we show that RPN4 is required for normal levels of these subunits. Further, we demonstrate that RPN4 is extremely short-lived ($t_{1/2}$ ≈2 min), that it directly interacts with RPN2, a subunit of the 26S proteasome, and that *rpn4Δ* cells are perturbed in their cell cycle. The degradation signal of RPN4 was mapped to its N-terminal region, outside the transcription–activation domains of RPN4. The ability of RPN4 to augment the synthesis of proteasomal subunits while being metabolically unstable yields a negative feedback circuit in which the same protein up-regulates the proteasome production and is destroyed by the assembled active proteasome.

proteolysis | ubiquitin | N-end rule | UFD pathway | cell cycle

The *Saccharomyces cerevisiae* RPN4 gene (its earlier names are *SON1* and *UFD5*) (1) was originally identified through mutant *rpn4* alleles that suppressed the growth defect of *sec63–101* cells, which bore a temperature-sensitive (ts) variant of SEC63, an essential component of the protein translocation channel in the endoplasmic reticulum membrane (2). More recent studies have shown that mutations in *RPN4* inhibit the degradation of normally short-lived proteins that are targeted by the N-end rule pathway, by the *u*biquitin/*f*usion/*d*egradation (UFD) pathway, and apparently also by other pathways of the ubiquitin (Ub)–proteasome system (3, 4). These findings suggested that the ability of *rpn4* mutations to suppress the conditional lethality of *sec63–101* may stem from stabilization of the mutant but partially active SEC63–101 against degradation at nonpermissive temperature.

Regulated proteolysis by the Ub/proteasome system plays essential roles in the cell cycle, differentiation, stress responses, and many other processes (5–7). Ub is a 76-residue protein whose covalent conjugation to other proteins marks these proteins for degradation by the 26S proteasome, an ATP-dependent multisubunit protease. Ub conjugation involves the formation of a thioester between the C terminus of Ub and a specific cysteine of the Ub-activating (E1) enzyme. The Ub moiety of E1~Ub thioester is transesterified to a cysteine in one of several Ub-conjugating (E2) enzymes. The Ub moiety of E2~Ub thioester is conjugated via the isopeptide bond to the ε-amino group of either a substrate's Lys residue or a Lys residue of another Ub moiety, the latter reaction resulting in a substrate-linked multi-Ub chain (7, 8). Most E2 enzymes function in complexes with proteins called E3 (9–11). The functions of E3s include the initial recognition of degradation signals (degrons) in substrate proteins, with different E3s recognizing different classes of degrons (12–14). The E2–E3 complexes, referred to as Ub ligases (this term is also used to denote E3s alone), mediate the formation of substrate-linked multi-Ub chains (15, 16). Ubiquitylated substrates are processively degraded by the 26S proteasome, which consists of the 20S core proteasome and two 19S particles (17–19). *In vivo*, the 20S proteasome exists in complexes with either the 19S particle or the 11S particle (of a distinct protein composition). The latter particle stimulates the peptidase but not the protease activity of the 20S proteasome (18). One 19S and one 11S particle can be bound to each end of the same 20S proteasome, a configuration of likely physiological significance (20, 21). The 19S particle mediates the binding and ATP-dependent unfolding of a substrate protein before its transfer to the interior of the 20S core (17). The biogenesis of the 20S proteasome has been analyzed in some detail (ref. 22 and refs. therein). Most, if not all, of the genes encoding the stoichiometrically present subunits of the *S. cerevisiae* 26S proteasome have been identified (1, 23–26), but regulation of these genes remains to be understood.

Consistent with the effects of *rpn4* mutations on Ub/proteasome-dependent proteolysis, RPN4 was reported to cofractionate with a partially purified 26S proteasome (27). However, RPN4 was not detected among proteasomal subunits in other analyses of purified 26S proteasomes (23, 25). Recent work identified a specific sequence motif in the promoters of yeast proteasomal genes and demonstrated that RPN4 binds to this motif and functions as a transcriptional activator of the motif-containing promoters (28).

In the present study, we showed that RPN4 is required for normal levels of proteasomal subunits in the cell. Further, we found that RPN4 is an extremely short-lived protein, that it directly interacts with at least one specific subunit of the 26S proteasome, and that cells lacking RPN4 are perturbed in their progression through post-G_1 phases of the cell cycle. We also characterized the degron of RPN4, locating it outside the putative transcription-activating domains. Our findings indicate that the ability of RPN4 to augment the synthesis of proteasomal subunits while being metabolically unstable yields a negative feedback circuit in which the intracellular proteolysis is up-regulated by a protein that is destroyed by the assembled active proteasome.

Materials and Methods

Strains, Plasmids, and β-Galactosidase (βgal) Assay. The *S. cerevisiae* strains used were EJY140 (*MATa trp1-Δ63 ura3-52 his3-Δ200 leu2-3, 112 lys2-801 rpn4Δ::LEU2*); JD52 (*MATa trp1-Δ63 ura3-52 his3-Δ200 leu2-3, 112 lys2-801*); AVY302 (*rpn2Δ::URA3* derivative of JD52); Y791 (*MATa cim5-1 ura3-52 his3-Δ200 leu2Δ1*); MHY501 (*MATa trp1-1 ura3-52 his3-Δ200 leu2-3, 112 lys2-801*); MHY1409 (*MATa trp1-1 ura3-52 his3-Δ200 leu2-3, 112 lys2-801 uba1-2*) (3, 29, 37). *RPN1* containing its promoter region was isolated as a suppressor of the toxicity

---

Abbreviations: ts, temperature sensitive; Ub, ubiquitin; βgal, *E. coli* β-galactosidase; UFD, *u*b/*f*usion/*d*egradation.

*To whom reprint requests should be addressed at: Division of Biology, 147-75, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125. E-mail: avarsh@caltech.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

www.pnas.org/cgi/doi/10.1073/pnas.071022298

of overexpressed N-end rule pathway (30). *PRE6* (containing its promoter region) and the ORF of *RPN4* were amplified by PCR from total DNA of *S. cerevisiae* YPH500. All constructs were verified by DNA sequencing. *RPN1* and *PRE6* (bearing their promoter regions) were subcloned into the low-copy vector pRS313 (31), yielding p313RPN1 and p313PRE6. The *RPN4* ORF was subcloned into the low-copy vector pRS314CUP1 derived from pRS314 (31), yielding p314CUP1RPN4, in which *RPN4* was expressed from the P_{CUP1} promoter. For immunoblotting and immunoprecipitation, the flag epitope was added to the C termini of RPN1, PRE6, and RPN4. Alternatively, the N terminus of RPN4 was extended with the ha epitope (32). For GST-pulldown assays, the ORFs of *RPN1*, *RPN2*, *RPN9*, *RPN10*, and *RPN12* were fused in frame to the 3′-end of the GST-coding sequence in pGEX-4T-3 (Amersham Pharmacia). The C-terminally flag-tagged derivative of *RPN4* was subcloned into pET-11c (Novagen). *Escherichia coli* BL21 (DE3) was used to express GST fusions, as well as RPN4-flag. The *E. coli lacZ* gene encoding βgal lacking the first eight residues was amplified by PCR, by using pMC1871 (Amersham Pharmacia) as a template. This βgal was expressed as a fusion to the C terminus of $RPN4_{1-151}$ from the P_{CUP1} promoter and pRS315 vector (31). $Ub^{K48R,G76A}$ was expressed from the P_{CUP1} promoter and a high-copy vector (33). Arg-βgal, derived from Ub-Arg-βgal, was expressed from the P_{GAL1} promoter and a high-copy vector (3). The activity of βgal in yeast extracts, from cultures at A_{600} of 0.8–1.0, was determined as described (32).

Immunoblotting, Pulse–Chase, and GST-Pulldown Assays. *S. cerevisiae* transformants were grown to OD_{600} of 0.8–1.0, harvested, and resuspended in lysis buffer (1% Triton X-100/0.15 M NaCl/1 mM EDTA/50 mM Na-Hepes, pH 7.5) containing 1× protease inhibitor mix (Boehringer Mannheim), and lysed by vortexing with glass beads. Equal amounts of extracts were separated by SDS/PAGE, followed by immunoblotting with monoclonal anti-flag antibody (Sigma) or anti-ha antibody (Covance, Berkeley, CA). SDS/PAGE in 6, 8, and 12% gels was used with, respectively, RPN1-flag, RPN4-flag or ha-RPN4, and PRE6-flag or truncated RPN4 proteins. Pulse–chase procedures were as described (30). Briefly, 10-ml cultures (OD_{600} of 0.8–1.2) of wild-type, *rpn2*Δ, *cim5-1*, and *uba1-2 S. cerevisiae* in SD media (30) containing 0.2 mM $CuSO_4$ were labeled for 5 min with 0.15 mCi of [^{35}S]methionine/cysteine (EXPRESS, New England Nuclear). Labeled cells were pelleted, resuspended in 0.8 ml of lysis buffer, and lysed as above. The extracts were centrifuged at 12,000 × *g* for 10 min, and supernatants containing equal amounts of CCl_3COOH-insoluble ^{35}S were used for immunoprecipitation with anti-flag, anti-ha, or anti-βgal antibodies (Promega). For binding assays with GST fusions (GST-pulldown assays), see ref. 30 and the legend to Fig. 1*E*.

Cell Cycle Analysis. Flow cytometric DNA analysis was performed as described previously (34). Briefly, exponentially growing cells (OD_{600} ≈1.0) were fixed in 70% ethanol and treated with RNase A (2 mg/ml) at 37°C for 2 h. Cells were then stained with propidium iodide (50 μg/ml) and analyzed by using Becton Dickinson FACScan. Cells ($\approx 2 \times 10^4$) were analyzed in each sample. For analyses of synchronous cultures, cells were grown to OD_{600} of ≈1.0 and treated with α factor (5 μg/ml) for 2 h, then washed and resuspended in yeast/peptone/dextrose (YPD medium). Light microscopic determination of the fraction of budded cells was carried out with samples taken after 2-h G_1 arrest and at 20-min intervals after release from arrest, with ≈200 cells analyzed from each sample. Synchronized cultures were also characterized by flow cytometry.

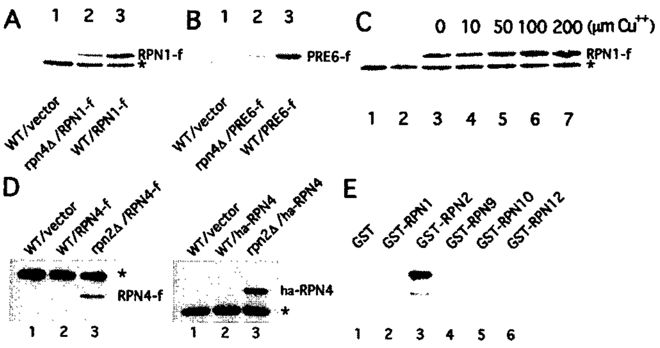


Fig. 1. RPN4 is required for normal expression of proteasomal subunits. (*A* and *B*) Immunoblot analysis of RPN1 (*A*) and PRE6 (*B*) that were C-terminally tagged with flag epitope and expressed from their native promoters and a low-copy plasmid in a *rpn4*Δ *S. cerevisiae* strain (lane 2) and its congenic wild-type counterpart JD52 (lane 3). Lane 1 in *A* and *B*, cells transformed with empty vector. (*C*) Enhanced expression of *RPN4* marginally elevates the level of RPN1. *rpn4*Δ cells were cotransformed with low-copy plasmids expressing, respectively, RPN1-flag and RPN4 from the P_{CUP1} promoter. Increasing concentrations of $CuSO_4$ were used to induce the expression of RPN4 (lanes 3–7). Lanes 1 and 2 in C, *rpn4*Δ cells transformed, respectively, with empty RPN1-flag vector and empty RPN4 vector. (*D*) Immunoblot analyses of C-terminally (RPN4-flag) or N-terminally (ha-RPN4) tagged RPN4 that was expressed from the induced P_{CUP1} promoter and low-copy plasmid either in wild-type (*RPN4*) strain JD52 or in AVY302, a congenic *rpn2*Δ mutant (see legend to Fig. 2 for details). (*E*) RPN4 interacts with RPN2 in GST-pulldown assays. Extracts of *E. coli* expressing RPN4-flag were incubated with glutathione-agarose beads preloaded with the indicated GST fusions or GST alone. The retained proteins were eluted, fractionated, and immunoblotted with anti-FLAG antibody. Approximately equal amounts of different GST fusions were immobilized on glutathione-agarose beads in these assays, as verified by Coomassie staining (data not shown). SDS/PAGE in 6, 8, and 12% gels was used, respectively, in *A* and *C*, in *D* and *E*, and in *B*. The asterisk in *A*, *C*, and *D* indicates a crossreacting band.

Results

RPN4 Is Required for Normal Expression of Proteasome Components. We initially attempted to identify *S. cerevisiae* RPN4-binding proteins by using the yeast two-hybrid assay and observed that RPN4 functioned as a transcriptional activator when fused to the DNA-binding domain of GAL4 (data not shown). This result was consistent with the finding that RPN4 binds to a DNA sequence motif present in the promoters of most genes encoding proteasomal subunits and several other genes of the Ub/proteasome system (28). We then examined the levels of two 26S proteasomal subunits, RPN1 (of the 19S particle) and PRE6 (of the 20S core proteasome) in the presence and absence of RPN4, by using immunoblotting (Fig. 1 *A* and *B*). C-terminally epitope-tagged RPN1-flag and PRE6-flag were expressed from their own promoters on a low-copy vector. The levels of RPN1-flag and PRE6-flag were significantly lower in a *rpn4*Δ strain than in a congenic wild-type strain (Fig. 1 *A* and *B*, lanes 2 and 3). Taken together with the gene expression data (28), these results indicated that RPN4 is a positive transcriptional regulator of genes encoding proteasomal subunits.

RPN4 Is a Short-Lived Protein Degraded by the 26S Proteasome. RPN4 was expressed from the copper-inducible P_{CUP1} promoter on a low-copy vector in the *rpn4*Δ strain that also expressed RPN1-flag from its natural promoter. The expression of *RPN4* from uninduced P_{CUP1} was sufficient to greatly augment the expression of RPN1-flag, but further enhancement of *RPN4* expression elevated the level of RPN1-flag only slightly (Fig. 1*C*, lanes 3–7). In agreement with this result, a strongly increased expression of *RPN4* in the wild-type (*RPN4*) background resulted in at most a slight enhancement of proteasome activity, as

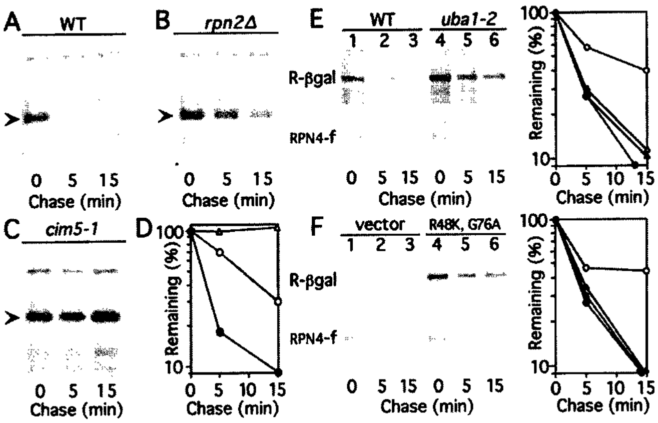


Fig. 2. RPN4 is a short-lived protein degraded by a proteasome-dependent pathway. (*A–C*) Pulse–chase analysis of C-terminally tagged RPN4 (RPN4-flag) that was expressed from the induced P_{CUP1} promoter and low-copy plasmid either in wild-type (*RPN4*) strain JD52 (*A*), in a congenic *rpn2Δ* mutant (*B*), or in a *cim5–1* mutant (*C*). Cells were labeled at 28°C and chased at 28°C in *A* and *B*, and at 37°C in *C*. The kinetics of RPN4 degradation in JD52 were similar at 28 and 37°C (data not shown). Arrowheads indicate the band of RPN4-flag. (*D*) Quantitation of the patterns in *A–C*, by using PhosphorImager (Molecular Dynamics) ●, wild-type cells. ○, *rpn2Δ* cells. △, *cim5–1* cells. (*E*) Pulse–chase analyses of Arg-βgal (R-βgal), derived from Ub-Arg-βgal (3, 14), and of RPN4-flag (expressed as in *A*) in *uba1–2* (37) and wild-type cells. Quantitation: ● and ○, Arg-βgal in wild-type and *uba1–2* cells, respectively. ▲ and △, RPN4-flag in the same strains. (*F*) The same test proteins in wild-type cells and cells overexpressing $Ub^{K48R, G76A}$, with quantitation on the right; same designations. The *rpn2Δ* locus, in the strain AVY302 (see *Materials and Methods*), was a disruption allele. It was produced through the integration of *URA3* at the *Hin*dIII site of *RPN2* (codon 146) and was identical to the *rpn2::URA3* allele described by Yokota *et al.* (36). The phenotypes of AVY302 and the previously described strain (36) were similar as well (data not shown). Two other studies reported that *rpn2Δ* cells were inviable (4, 56). The disruption of *RPN2* in these works was carried out with either *TRP1* or *ADE2*, by using the *RPN2 Bgl*II site at codon 38. It remains to be determined whether different integration sites account for different phenotypes described for *rpn2Δ* strains.

assessed by measuring the steady-state levels of test substrates that are targeted by either the N-end rule pathway or the UFD pathway (data not shown), both pathways being Ub/proteasome-dependent (3, 35). By contrast, expression of *RPN4* in *rpn4Δ* cells greatly increased the activity of these proteolytic pathways (data not shown).

Given these results, we expressed either a C-terminally tagged RPN4-flag or an N-terminally tagged ha-RPN4 from the P_{CUP1} promoter and determined the levels of RPN4 protein at different concentrations of the promoter-inducing $CuSO_4$. Neither of these tagged derivatives of RPN4 could be detected by immunoblotting even on the induction of P_{CUP1} (Fig. 1*D*), suggesting that RPN4 was a short-lived protein, a property that could account for the relative insensitivity of RPN4-dependent promoters to the level of transcriptional activity of an RPN4-expressing gene. Pulse–chase assays were then carried out with RPN4-flag in wild-type (*RPN4*) cells, revealing that the *in vivo* half-life of RPN4-flag was $\approx$2 min (Fig. 2*A* and *D*). The half-life of N-terminally tagged ha-RPN4 was indistinguishable from that of C-terminally tagged RPN4-flag (data not shown), ruling out the epitope tags as the cause of RPN4 metabolic instability. *S. cerevisiae* cells lacking RPN2, a protein of the 19S component of the 26S proteasome, are partially defective in the degradation of natural substrates of the proteasome (36) (see also the legend to Fig. 2). Both RPN4-flag and ha-RPN4 could be detected by immunoblotting in *rpn2Δ* cells, in contrast to wild-type (*RPN2*) cells (Fig. 1*D*). Consistent with these results, pulse–chase assays indicated a significant decrease in degradation of RPN4-flag in *rpn2Δ* cells, in comparison to wild-type cells (Fig. 2 *B* and *D*). In addition, RPN4-flag was found to be long-lived in the *cim5–1* mutant (Fig. 2 *C* and *D*), which bears a ts mutation in RPT1, an essential ATPase of the 19S particle (29). We conclude that the *in vivo* degradation of RPN4 is mediated by the 26S proteasome.

RPN4 Interacts with RPN2, a Component of the 19S Particle. It was reported that RPN4 cofractionated with the 26S proteasome (27), but RPN4 was not observed in other studies of purified 26S proteasomes (23, 26). Previous work (30) used the GST-pulldown assay to demonstrate that UBR1 and UFD4, the Ub ligase (E3) components of, respectively, the N-end rule and UFD pathways, directly interact with specific subunits of the proteasome's 19S particle. We used the GST-pulldown approach to determine whether RPN4 might interact with specific subunits of the 19S particle. In these experiments, several proteins of the 19S particle were expressed in *E. coli* as fusions to the C terminus of GST. Extracts from *E. coli* expressing RPN4-flag were incubated with glutathione-agarose beads preloaded with GST-RPN1, GST-RPN2, GST-RPN9, GST-RPN10, GST-RPN12, or GST alone. The bound proteins were eluted, fractionated by SDS/PAGE, and immunoblotted with anti-flag antibody. RPN4-flag reproducibly bound to GST-RPN2 but not to any of the other tested subunits of the 19S particle (Fig. 1*E*). It remains to be determined whether RPN2 is the only proteasomal ligand of RPN4, or whether some other proteasomal subunits, among the still untested ones, also interact with RPN4.

Because the steady-state level of RPN4 in wild-type cells was sufficiently low to be undetectable by standard immunoblotting (Fig. 1*D*), RPN4 cannot be a stoichiometric component of the 19S particle. Furthermore, because RPN4 is also a transcriptional regulator (28), it is clear that the population of RPN4 molecules in a cell is dynamically partitioned among several classes of physiologically relevant complexes that include the 26S proteasome and either specific or nonspecific RPN4-binding sites on the chromosomes.

Is Degradation of RPN4 Ubiquitin-Dependent? As demonstrated above (Figs. 1*D* and 2), RPN4 is degraded by the 26S proteasome. Several lines of evidence suggested that ubiquitylation plays at most a minor role in the proteasome-dependent degradation of RPN4. First, the $\approx$2-min half-life of RPN4 in wild-type cells (Fig. 2*D*) was not significantly changed in the *uba1–2* mutant, which underexpresses the Ub-activating (E1) enzyme and is therefore strongly impaired in ubiquitylation of proteins (37). Specifically, whereas the degradation of Arg-βgal, a substrate of the Ub-dependent N-end rule pathway (14, 38), was decreased in the *uba1–2* mutant, no significant change was observed with RPN4-flag in this mutant (Fig. 2*E*). Second, overexpression of $Ub^{K48R,G76A}$, a Ub mutant that inhibits the formation of Lys^{48}-linked multi-Ub chains, which are essential for a large fraction of the proteasome-dependent proteolysis (33, 39), did not significantly decrease the rate of RPN4 degradation but did decrease the degradation of Arg-βgal (Fig. 2*F*). Third, the kinetics of degradation of RPN4-flag in wild-type cells were indistinguishable from that in mutants lacking one of the following Ub-conjugating (E2) enzymes: UBC1, RAD6 (UBC2), CDC34 (UBC3), UBC4, UBC5, UBC6, UBC7, or UBC8 (data not shown). Analogous pulse–chase assays were also carried out with mutants lacking different pairs of these E2 enzymes. The degradation of RPN4 was slightly decreased in [*ubc4Δ ubc5Δ*] cells, which lacked UBC4 and UBC5, two highly similar E2s (data not shown). Because UBC4/UBC5 are a functionally major class of E2 enzymes in *S. cerevisiae* (7), and because [*ubc4Δ ubc5Δ*] cells grow slowly and exhibit a number of defects, the observed marginal stabilization of RPN4 in a [*ubc4Δ ubc5Δ*] strain could be an indirect result of multiple changes that are caused by the absence of UBC4/UBC5. Fourth, the degradation of RPN4 was not decreased in *S. cerevisiae* mutants that lacked

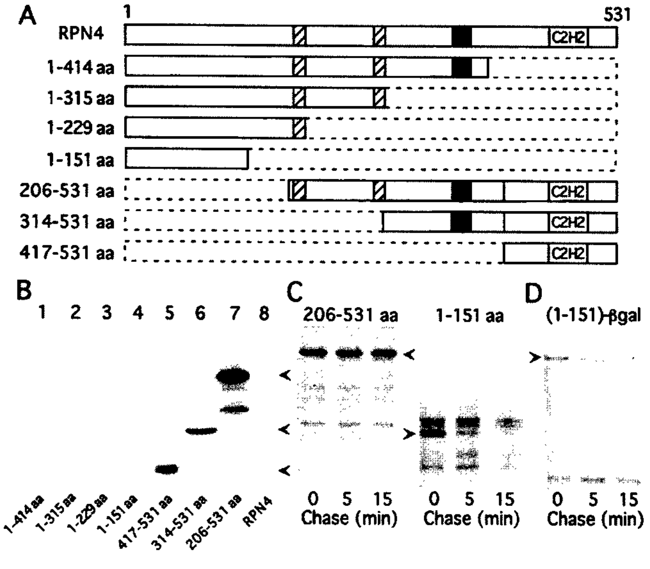


Fig. 3. The N-terminal region of RPN4 contains a portable degradation signal. (*A*) Diagrams of full-length and truncated RPN4. Two putative transcription activation domains are at positions 211–229 and 300–315 (hatched boxes). A putative bipartite nuclear localization signal is at position 381–399 (black box). A putative C_2H_2 finger (residues 477–507) is also indicated. (*B*) N-terminally truncated RPN4 derivatives are long-lived. RPN4 and its truncated derivatives were expressed from the P_{CUP1} promoter and low-copy vector in the JD52 (*RPN4*) strain. Relative steady-state levels of truncated RPN4 proteins (all of them tagged C-terminally with flag epitope) were determined by immunoblotting, by using SDS/12% PAGE and anti-flag antibody. The bands of proteins with expected sizes are indicated by arrowheads. (*C*) RPN4$_{1-151}$-flag is short-lived. Pulse–chase assays, by using SDS/12% PAGE, were carried out in wild-type (*RPN4*) cells with RPN4$_{206-531}$-flag and RPN4$_{1-151}$-flag (indicated by arrowheads) essentially as described in Fig. 2, except that they were performed at 30°C instead of 28°C. (*D*) The 151-residue N-terminal fragment of RPN4 contains a portable degron. Pulse–chase assay with RPN4$_{1-151}$-βgal fusion (see *Materials and Methods*) expressed from the P_{CUP1} promoter and low-copy vector in the JD52 (*RPN4*) strain, by using SDS/6% PAGE and immunoprecipitation with anti-βgal antibody.

one of the following E3 enzymes: UBR1, UFD4, RSP5, TOM1, HUL4, or HUL5 (data not shown) (3, 14, 40, 41). Thus, it appears, but remains to be established definitively, that the proteasome-dependent degradation of RPN4 (Fig. 2 *A* and *D*) is either largely or completely independent of RPN4 ubiquitylation.

N-Terminal Region of RPN4 Contains a Portable Degron. The degradation signals of transcriptional activators are often located within their activation domains (42–45). We constructed a set of truncated, C-terminally tagged RPN4 derivatives that contained either two, one, or none of the putative transcription-activation domains of the intact RPN4 (Fig. 3*A*). Whereas wild-type RPN4 was undetectable by immunoblotting, because of its rapid *in vivo* degradation (Figs. 1*D* and 2 *A* and *D*), RPN4 derivatives such as RPN4$_{206-531}$, RPN4$_{314-531}$, and RPN4$_{417-531}$, which lacked at least the first 205 residues of RPN4, were readily detectable (Fig. 3*B*, lanes 5–8). By contrast, similarly expressed truncated derivatives that retained this N-terminal region of RPN4 (RPN4$_{1-414}$, RPN4$_{1-315}$, RPN4$_{1-229}$, and RPN4$_{1-151}$) were still largely undetectable by immunoblotting (Fig. 3*B*, lanes 1–4), consistent with the interpretation that the major degron of RPN4 was located within its first 150–200 residues. This region does not contain putative transcription-activation domains (Fig. 3*A*).

Pulse–chase assays with RPN4$_{1-151}$-flag (containing the presumed degron of RPN4) and RPN4$_{206-531}$-flag (lacking this degron) confirmed the inferences from immunoblotting data: RPN4$_{1-151}$-flag was rapidly degraded ($t_{1/2}$ ≈2 min), whereas RPN4$_{206-531}$-flag was long-lived (Fig. 3*C*). The degron of RPN4 is portable, in that ligation of RPN4$_{1-151}$ to the otherwise long-lived 115-kDa βgal moiety resulted in a metabolically unstable RPN4$_{1-151}$-βgal protein ($t_{1/2}$ ≈5 min) (Fig. 3*D*). No sequence similarities between the 151-residue region of RPN4 and the known motifs that act as degrons in other short-lived proteins (13) could be detected, suggesting that this region of RPN4 contains a degradation signal which, to our knowledge, is novel. Interestingly, the C-terminally truncated, degron-containing, metabolically unstable RPN4$_{1-229}$ did not bind to RPN2 of the 19S particle in the GST-pulldown assay (data not shown), in contrast to full-length RPN4 (Fig. 1*E*), suggesting that the *in vivo* degradation of intact RPN4 ($t_{1/2}$ ≈2 min) may be independent of the demonstrated RPN4–RPN2 interaction.

Cells Lacking RPN4 Exhibit Delay in Post-G$_1$ Phases of the Cell Cycle. Despite the decreased expression of proteasomal components in *rpn4*Δ cells (Fig. 1), they were not only viable but also similar to congenic wild-type cells in their resistance to ≈260-nm UV light and in their ability to grow at high temperature (37°C), on poor nitrogen sources (proline and citrulline), and in the presence of canavanine (a toxic arginine analog, at 1.5 μg/ml), NaCl (at 1 M), $CuSO_4$ (at 0.5 mM), or ethanol (at 3%) (data not shown). At the same time, *rpn4*Δ cells are known to grow more slowly at 30°C on standard media than wild-type cells [(2–4); unpublished data]. To assess the effect of RPN4 absence on cell cycle progression, we carried out flow cytometric analyses of DNA content with unsynchronized exponentially growing cultures of haploid *rpn4*Δ and congenic wild-type cells. The fraction of cells with 2*N* (replicated) DNA content was significantly higher in *rpn4*Δ culture (Fig. 4*A*). Light microscopic examination indicated a significantly higher fraction of budded (post-G$_1$) cells in exponentially growing *rpn4*Δ cultures, in comparison to wild-type ones (data not shown), suggesting that cells lacking RPN4 are delayed in their progression through one or more of post-G$_1$ phases of the cell cycle.

To address this question in greater detail, *rpn4*Δ and congenic wild-type cultures were synchronized in G$_1$ phase with α factor, followed by flow cytometric analysis of cellular DNA content as a function of time after release from G$_1$ arrest. As shown in Fig. 4*B*, *rpn4*Δ cells exited G$_1$ phase without delay relative to wild-type cells. However, it took longer for *rpn4*Δ cells to progress through the rest of the cell cycle (Fig. 4*B*). Light microscopy was used to determine the fraction of budded (post-G$_1$) cells in synchronized cultures at different times after resumption of growth. Consistent with the results of flow cytometric analysis, the budding of *rpn4*Δ and wild-type cells occurred at approximately the same time (peak at ≈80 min) (Fig. 4*C*). However, it took longer, on average, for the budded *rpn4*Δ cells to complete cell division (≈140 min for *rpn4*Δ cells versus ≈120 min for wild-type cells). Note also that a smaller fraction of budded cells exited from M to G$_1$ in the *rpn4*Δ culture than in the wild-type culture (Fig. 4*C*), consistent with the data for unsynchronized cultures (Fig. 4*A*). Because the ratio of small- to large-budded cells was essentially the same for the synchronized wild-type and *rpn4*Δ cultures at different time points (data not shown), it is likely that the absence of RPN4 affects more than one post-G$_1$ step, e.g., both the S → G$_2$ and G$_2$ → M transitions.

Discussion

RPN4 was originally identified as an extragenic suppressor of *sec63-101*, a ts allele of *SEC63*, which is required for the translocation of proteins into the endoplasmic reticulum and the transport of proteins to the nucleus (2). *rpn4* mutations suppressed the ts growth defect of *sec63-101* but did not reverse its phenotype of mislocalization of proteins that bore a nuclear targeting signal. Moreover, *rpn4* mutants themselves had a

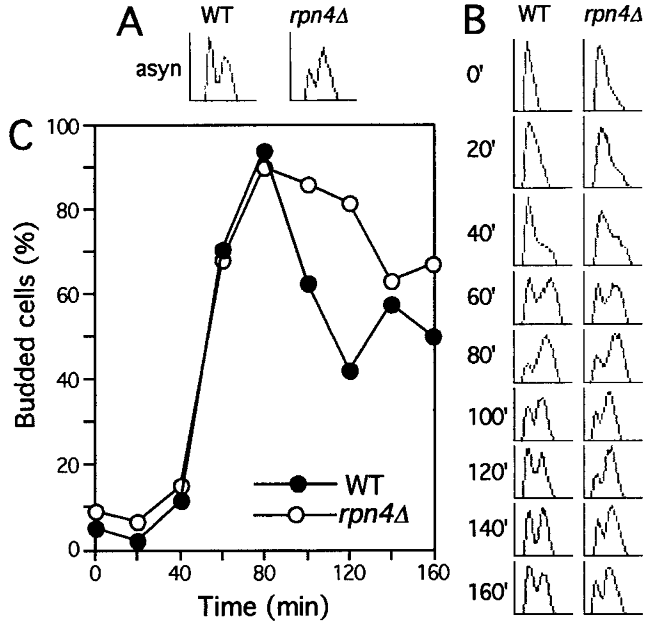


Fig. 4. Cell cycle progression is delayed after G_1 phase in *rpn4*Δ mutant. (*A*) A larger fraction of *rpn4*Δ cells have 2*N* DNA content in unsynchronized cultures. DNA flow cytometry was carried out with wild-type (*RPN4*) and *rpn4*Δ cells. The abscissa ordinate indicate, respectively, the number of cells and relative DNA content. (*B* and *C*) Cell cycle progression of *rpn4*Δ cells is delayed after G_1 phase. Wild-type and congenic *rpn4*Δ cells were synchronized in late G_1 by using α factor. Initial samples were taken after 2 h of G_1 arrest, and subsequent samples were taken at 20-min intervals after release from the arrest. The samples were analyzed by using DNA flow cytometry (*B*) and by determining the fraction of budded cells (*C*). Data shown in *C* are the means of triplicate measurements.

defect in the nuclear protein import (2). Later work has shown that mutations in *RPN4* inhibit the activity of the Ub/proteasome-dependent N-end rule and UFD proteolytic pathways (3). Our findings (Fig. 1 *A* and *B*) and a recent study (28) demonstrated that RPN4 is required for normal expression of subunits of the 26S proteasome and several other components of the Ub system. A parsimonious interpretation is that the observed down-regulation of the Ub/proteasome-dependent proteolysis in *rpn4*Δ cells results primarily from decreased expression of proteasomal subunits and functionally related proteins.

Previous studies did not reach a consensus on whether RPN4 is a component of the 26S proteasome (23, 24, 26, 27). A direct and specific interaction between RPN2 (a stoichiometric subunit of the 19S particle) and RPN4 was demonstrated in the present work (Fig. 1*E*). This result, together with the earlier finding that the E3 enzymes UBR1 and UFD4 interact with specific subunits of the 19S particle (30) and the evidence for interactions between other E2/E3 enzymes and the 26S proteasome (23, 46), indicates that proteasome is at least a transient ligand of many cellular proteins.

An operationally useful definition of a bona fide subunit of the 26S proteasome should stipulate that a large (predetermined) fraction of the proteasome particles in a cell is associated with this subunit *in vivo*. By contrast, and similarly to terminology in the ribosome field, specific protein ligands of the proteasome that interact with it transiently and/or are bound to small, dynamically determined subsets of the 26S proteasome, can be called proteasome-interacting proteins. In this terminology, RPN4, UBR1, and UFD4 are proteasome-interacting proteins, as distinguished from the 26S proteasomal subunits such as PRE6, RPT6, or RPN2.

RPN4 is not essential for cell viability under normal conditions. As described above, *rpn4*Δ cells are not hypersensitive to a variety of physical and chemical stresses, suggesting that a strongly decreased concentration of the 26S proteasome in *rpn4*Δ cells is sufficient to maintain cell viability and growth even under conditions of stress. At the same time, given the diminished activity of at least the N-end rule and UFD pathways in *rpn4*Δ cells (3), one would expect an impairment of some physiological functions that require these pathways. The post-G_1 abnormality in the cell cycle progression of *rpn4*Δ cells described in the present work (Fig. 4) is unlikely to be caused by inhibition of the N-end rule pathway, because *ubr1*Δ cells, which lack this pathway, lose chromosomes at a greatly increased frequency but are similar to congenic wild-type cells in the kinetics of cell cycle progression (47). [The chromosome-loss phenotype of *ubr1*Δ cells is caused largely by metabolic stabilization of the ESP1-produced fragment of SCC1, a prosubstrate of the N-end rule pathway and a component of the chromosome-bound cohesin complex (47).] Because RPN4 functions as a transcriptional activator of many nonproteasomal genes as well (48), the cell cycle defect of *rpn4*Δ cells may stem from down-regulation of these other genes. For example, the expression of *CDC48*, which is essential for the cell cycle progression and encodes an AAA-type ATPase, is decreased in *rpn4*Δ cells (28). However, we found that overexpression of *CDC48* from a heterologous (P_{MET25}) promoter did not rescue the cell cycle abnormality of *rpn4*Δ cells (data not shown).

One of our main results is the striking metabolic instability of RPN4 ($t_{1/2} \approx 2$ min). The degradation of RPN4 is proteasome-dependent (Fig. 2 *A*–*D*). At the same time, several lines of evidence (Fig. 2 *E* and *F*; see also above) suggest that degradation of RPN4 is largely independent of ubiquitylation. Proteins whose *in vivo* degradation is proteasome-dependent but Ub-independent include ornithine decarboxylase (49) and the cyclin-dependent kinase inhibitor $p21^{Cip1}$ (50, 51).

The mapping of RPN4 degron localized it to the first 150 residues of the 531-residue RPN4 (Fig. 3), outside of its putative transcription activation domains. This location of a degron is an exception to the previously established pattern in which the activation domains and degrons tend to overlap in a transcriptional regulator (42–45). As to the mechanics of RPN4 degron, the properties of RPN4 suggest two possibilities. The N-terminal degradation signal of RPN4 may function as a canonical bipartite degron of the Ub/proteasome system (12, 13, 52). The first determinant of such a degron is bound by a degron-specific E2-E3 Ub ligase, whereas the second determinant is an internal Lys residue of a substrate protein. Alternatively, RPN4 might be targeted for degradation directly through its demonstrated interaction with RPN2, a subunit of the 19S particle (Fig. 1*E*). This model is made unlikely (but not definitively precluded) by the fact that the degron-containing, short-lived $RPN4_{1-229}$ fragment was unable to bind to RPN2, in contrast to full-length RPN4 (see above).

Metabolic instability of RPN4 identifies it as a member of a growing class of components of the Ub/proteasome system that are also substrates of this system and are destroyed by it, either constitutively or conditionally. One example is UMP1, a chaperone that is required for efficient assembly of the 20S proteasome and becomes its first substrate on the completion of assembly (22, 53). Other examples are *S. cerevisiae* F-box proteins CDC4, GRR1, and MET30, which are short-lived substrate-recognition subunits of, respectively, SCF^{Cdc4}, SCF^{Grr1}, and SCF^{Met30} Ub ligases (54, 55).

The unique ability of RPN4 to augment the transcription of genes encoding proteasomal subunits while being metabolically unstable yields a negative feedback circuit in which the same protein up-regulates the proteasome production and is destroyed by the assembled active proteasome.

We are grateful to Xiaofeng Qin and Wenge Lu for advice and assistance with flow cytometric analysis, to Christopher Byrd (California Institute of Technology), Jürgen Dohmen (University of Cologne, Cologne, Germany), Mark Hochstrasser (Yale University, New Haven, CT), Jon Huibregtse (Rutgers University, Piscataway, NJ), Erica Johnson (Thomas Jefferson University, Philadelphia, PA), Yoshiko Kikuchi (University of Tokyo, Tokyo, Japan), Robert Swanson (University of Chicago, Chicago, IL), Glenn Turner (California Institute of Technology), and Fred Winston (Harvard Medical School, Boston, MA) for strains and constructs, and to Keiji Tanaka (Institute of Medical Science, Tokyo) for anti-RPN4 antibody. This work was supported by grants to A.V. from the National Institutes of Health (NIH) (DK39520 and GM31530). Y.X. was supported in part by a postdoctoral fellowship from the NIH.

1. Finley, D., Tanaka, K., Mann, C., Feldmann, H., Hochstrasser, M., Vierstra, R., Johnston, S., Hampton, R., Haber, J., McCusker, J., *et al* (1998) *Trends Biochem. Sci.* **23,** 244–245.
2. Nelson, M. K., Kurihara, T. & Silver, P. A. (1993) *Genetics* **134,** 159–173.
3. Johnson, E. S., Ma, P. C., Ota, I. M. & Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 17442–17456.
4. Xu, B.-E. & Kurjan, J. (1997) *Mol. Biol. Cell* **8,** 1649–1664.
5. Hershko, A., Ciechanover, A. & Varshavsky, A. (2000) *Nat. Med.* **10,** 1073–1081.
6. Varshavsky, A. (1997) *Trends Biochem. Sci.* **22,** 383–387.
7. Hochstrasser, M. (1996) *Annu. Rev. Genet.* **30,** 405–439.
8. Scheffner, M., Smith, S. & Jentsch, S. (1998) in *Ubiquitin and the Biology of the Cell,* eds. Peters, J.-M., Harris, J. R. & Finley, D. (Plenum, New York), pp. 65–98.
9. Deshaies, R. J. (1999) *Annu. Rev. Cell Dev. Biol.* **15,** 435–467.
10. Tyers, M. & Jorgensen, P. (2000) *Curr. Opin. Genet. Dev.* **10,** 54–64.
11. Schulman, B. A., Carrano, A. C., Jeffrey, P. D., Bowen, Z., Kinnucan, E. R. E., Finnin, M. S., Elledge, S. J., Harper, J. W., Pagano, M. & Pavletich, N. (2000) *Nature (London)* **408,** 381–386.
12. Dohmen, R. J. (2000) in *Proteasomes: The World of Regulatory Proteolysis,* eds. Hilt, W. & Wolf, D. (Landes, Georgetown, TX), pp. 188–205.
13. Laney, J. D. & Hochstrasser, M. (1999) *Cell* **97,** 427–430.
14. Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 12142–12149.
15. Pickart, C. M. (1997) *FASEB J.* **11,** 1055–1066.
16. Dubiel, W. & Gordon, C. (1999) *Curr. Biol.* **9,** 554–557.
17. Voges, D., Zwickl, P. & Baumeister, W. (1999) *Annu. Rev. Biochem.* **68,** 1015–1068.
18. DeMartino, G. N. & Slaughter, C. A. (1999) *J. Biol. Chem.* **274,** 22123–22126.
19. Rechsteiner, M. (1998) in *Ubiquitin and the Biology of the Cell,* eds. Peters, J. M., Harris, J. R. & Finley, D. (Plenum, New York), pp. 147–189.
20. Hendil, K. B., Khan, S. & Tanaka, K. (1998) *Biochem. J.* **332,** 749–754.
21. Whitby, F. G., Masters, E. I., Kramer, L., Knowlton, J. R., Yao, Y., Wang, C. C. & Hill, C. P. (2000) *Nature (London)* **408,** 115–120.
22. Ramos, P. C., Höckendorff, J., Johnson, E. S., Varshavsky, A. & Dohmen, R. J. (1998) *Cell* **92,** 489–499.
23. Verma, R., Chen, S., Feldman, R., Schieltz, D., Yates, J., Dohmen, R. J. & Deshaies, R. J. (2000) *Mol. Biol. Cell* **11,** 3425–3439.
24. Russell, S. J., Steger, K. A. & Johnston, S. A. (1999) *J. Biol. Chem.* **274,** 21943–21952.
25. Glickman, M. H., Rubin, D. M., Coux, O., Wefes, I., Pfeifer, G., Cjeka, Z., Baumeister, W., Fried, V. A. & Finley, D. (1998) *Cell* **94,** 615–623.
26. Glickman, M. H., Rubin, D. M., Fried, V. A. & Finley, D. (1998) *Mol. Cell. Biol.* **18,** 3149–3162.
27. Fujimoro, M., Tanaka, K., Yokosawa, H. & Toh-e, A. (1998) *FEBS Lett.* **423,** 149–154.
28. Mannhaupt, G., Schnall, R., Karpov, V., Vetter, I. & Feldmann, H. (1999) *FEBS Lett.* **450,** 27–34.
29. Ghislain, M., Udvardy, A. & Mann, C. (1993) *Nature (London)* **366,** 358–362.
30. Xie, Y. & Varshavsky, A. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 2497–2502. (First Published February 25, 2000; 10.1073/pnas.060025497)
31. Sikorski, R. S. & Hieter, P. (1989) *Genetics* **122,** 19–27.
32. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G. & Struhl, K. (1998) *Current Protocols in Molecular Biology* (Wiley, New York).
33. Finley, D., Sadis, S., Monia, B. P., Boucher, P., Ecker, D. J., Crooke, S. T. & Chau, V. (1994) *Mol. Cell. Biol.* **14,** 5501–5509.
34. Haase, S. & Lew, D. J. (1997) *Methods Enzymol.* **283,** 322–332.
35. Baker, R. T. & Varshavsky, A. (1991) *Proc. Natl. Acad. Sci. USA* **87,** 2374–2378.
36. Yokota, K., Kagawa, S., Shimizu, Y., Akioka, H., Tsurumi, C., Noda, C., Fujimuro, M., Yokosawa, H., Fujiwara, T., Takahashi, E., *et al.* (1996) *Mol. Biol. Cell* **7,** 853–870.
37. Swanson, R. & Hochstrasser, M. (2000) *FEBS Lett.* **477,** 193–198.
38. Kwon, Y. T., Balogh, S. A., Davydov, I. V., Kashina, A. S., Yoon, J. K., Xie, Y., Gaur, A., Hyde, L., Denenberg, V. H. & Varshavsky, A. (2000) *Mol. Cell. Biol.* **20,** 4135–4148.
39. Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K. & Varshavsky, A. (1989) *Science* **243,** 1576–1583.
40. Wang, G., Yang, J. & Huibregtse, J. M. (1999) *Mol. Cell. Biol.* **19,** 342–352.
41. Sasaki, T., Toh-e, A. & Kikuchi, Y. (2000) *Mol. Gen. Genet.* **262,** 940–948.
42. Flinn, E. M., Busch, C. M. C. & Wright, A. P. H. (1998) *Mol. Cell. Biol.* **18,** 5961–5969.
43. Molinari, E., Gilman, M. & Natesan, S. (1999) *EMBO J.* **22,** 6439–6447.
44. Wang, D., Moriggl, R., Stravopodis, D., Carpino, N., Marine, J.-C., Teglund, S., Feng, J. & Ihle, J. N. (2000) *EMBO J.* **19,** 392–399.
45. Salghetti, S. E., Muratani, M., Wijnen, H., Futcher, B. & Tansey, W. P. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 3118–3123. (First Published March 7, 2000; 10.1073/pnas.050007597)
46. Tongaonkar, P., Chen, L., Lambertson, D., Ko, B. & Madura, K. (2000) *Mol. Cell. Biol.* **20,** 4691–4698.
47. Rao, H., Uhlmann, F., Nasmyth, K. & Varshavsky, A. (2001) *Nature (London)* (in press).
48. Jelinsky, S. A., Estep, P., Church, G. M. & Samson, L. D. (2000) *Mol. Cell. Biol.* **20,** 8157–8167.
49. Coffino, P. (1998) in *Ubiquitin and the Biology of the Cell,* eds. Peters, J. M., Harris, J. R. & Finley, D. (Plenum, New York), pp. 411–428.
50. Sheaff, R. J., Singer, J. D., Swanger, J., Smitherman, M., Roberts, J. M. & Clurman, B. E. (2000) *Mol. Cell* **5,** 403–410.
51. Verma, R. & Deshaies, R. J. (2000) *Cell* **101,** 341–344.
52. Bachmair, A. & Varshavsky, A. (1989) *Cell* **56,** 1019–1032.
53. Burri, L., Höckendorff, J., Boehm, U., Klamp, T., Dohmen, R. J. & Lévy, F. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 10348–10353. (First Published September 5, 2000; 10.1073/pnas.190268597)
54. Zhou, P. & Howley, P. M. (1998) *Mol. Cell* **2,** 571–580.
55. Rouillon, A., Barbey, R., Patton, E., Tyers, M. & Thomas, D. (2000) *EMBO J.* **19,** 282–294.
56. DeMarini, D. J., Papa, F. R., Swaminathan, S., Ursic, D., Rasmussen, T. P., Culbertson, M. R. & Hochstrasser, M. (1995) *Mol. Cell. Biol.* **15,** 6311–6321.

MOLECULAR AND CELLULAR BIOLOGY, June 2000, p. 4135–4148
0270-7306/00/$04.00+0
Copyright © 2000, American Society for Microbiology. All Rights Reserved.

Vol. 20, No. 11

Altered Activity, Social Behavior, and Spatial Memory in Mice Lacking the NTAN1p Amidase and the Asparagine Branch of the N-End Rule Pathway

YONG TAE KWON,[1] SETH A. BALOGH,[2] ILIA V. DAVYDOV,[1] ANNA S. KASHINA,[1] JEONG KYO YOON,[1]
YOUMING XIE,[1] ARTI GAUR,[1]† LYNN HYDE,[2]‡ VICTOR H. DENENBERG,[2,3]
AND ALEXANDER VARSHAVSKY[1]*

Division of Biology, California Institute of Technology, Pasadena, California 91125,[1] and Biobehavioral Sciences Graduate Program[2] and Department of Psychology,[3] University of Connecticut, Storrs, Connecticut 06269-4154

Received 30 December 1999/Returned for modification 28 February 2000/Accepted 8 March 2000

The N-end rule relates the in vivo half-life of a protein to the identity of its N-terminal residue. N-terminal asparagine and glutamine are tertiary destabilizing residues, in that they are enzymatically deamidated to yield secondary destabilizing residues aspartate and glutamate, which are conjugated to arginine, a primary destabilizing residue. N-terminal arginine of a substrate protein is bound by the *Ubr1*-encoded E3α, the E3 component of the ubiquitin-proteasome-dependent N-end rule pathway. We describe the construction and analysis of mouse strains lacking the asparagine-specific N-terminal amidase (Nt^N-amidase), encoded by the *Ntan1* gene. In wild-type embryos, *Ntan1* was strongly expressed in the branchial arches and in the tail and limb buds. The *Ntan1*$^{-/-}$ mouse strains lacked the Nt^N-amidase activity but retained glutamine-specific Nt^Q-amidase, indicating that the two enzymes are encoded by different genes. Among the normally short-lived N-end rule substrates, only those bearing N-terminal asparagine became long-lived in *Ntan1*$^{-/-}$ fibroblasts. The *Ntan1*$^{-/-}$ mice were fertile and outwardly normal but differed from their congenic wild-type counterparts in spontaneous activity, spatial memory, and a socially conditioned exploratory phenotype that has not been previously described with other mouse strains.

A multitude of regulatory circuits involve conditionally or constitutively short-lived proteins (26, 27, 44, 48, 49, 64). Features of proteins that confer metabolic instability are called degradation signals, or degrons (37, 63). The essential component of one degradation signal, termed the N-degron, is a destabilizing N-terminal residue of a protein (3). A set of N-degrons containing different N-terminal residues which are destabilizing in a given cell yields a rule, termed the N-end rule, which relates the in vivo half-life of a protein to the identity of its N-terminal residue. An N-end rule pathway is present in all organisms examined, from mammals and plants to fungi and prokaryotes (63).

In eukaryotes, an N-degron comprises two determinants: a destabilizing N-terminal residue and an internal lysine of a substrate protein (4, 32, 60). The Lys residue is the site of formation of a substrate-linked multiubiquitin chain (15, 49). The N-end rule pathway is thus one pathway of the ubiquitin (Ub) system (25–27). Ub is a 76-residue eukaryotic protein that exists in cells either free or covalently conjugated to many other proteins. The Ub system plays a role in a vast range of processes, including cell growth, division, differentiation, and responses to stress. In most of these processes, Ub acts through routes that involve the degradation of Ub-protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease (10, 17, 20, 51).

(Throughout the text, the names of mouse genes are in italics, with the first letter uppercase. The names of human and *Saccharomyces cerevisiae* genes are also in italics, all uppercase. If human and mouse genes are named in the same sentence, the mouse gene notation is used. The names of *S. cerevisiae* proteins are roman, with the first letter uppercase and an extra lowercase "p" at the end. The names of mouse and human proteins are the same, except that all letters but the last "p" are uppercase. The latter usage is a modification of the existing convention [58], to facilitate simultaneous discussions of yeast, mouse, and human proteins. In some citations, the abbreviated name of a species precedes the gene's name.)

The N-end rule has a hierarchic structure. In the yeast *S. cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by the *NTA1*-encoded N-terminal amidohydrolase (Nt-amidase), into the secondary destabilizing N-terminal residues Asp and Glu (6). Destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *S. cerevisiae ATE1*-encoded Arg-tRNA protein transferase (R-transferase) (8, 41), to Arg, one of the primary destabilizing residues (Fig. 1A). In mammals, the deamidation step is mediated by two Nt-amidases, Nt^N-amidase and Nt^Q-amidase, which are specific, respectively, for N-terminal Asn and Gln (Fig. 1A) (24, 59). The mammalian counterpart of the yeast R-transferase Ate1p exists as two distinct species, ATE1-1p and ATE1p-2, which are produced through alternative splicing of *Ate1* pre-mRNA (34). In vertebrates, the set of secondary destabilizing residues contains not only Asp and Glu but also Cys, which is a stabilizing residue in yeast (Fig. 1A) (18, 23). The primary destabilizing N-terminal residues are bound directly by the *UBR1*-encoded N-recognin, the targeting (E3) component of the N-end rule pathway. In *S. cerevisiae*, Ubr1p is a 225-kDa protein which recognizes potential N-end rule substrates through its type 1 and type 2 sub-

---

* Corresponding author. Mailing address: Alexander Varshavsky, Division of Biology, 147-75, Caltech, 1200 East California Blvd., Pasadena, CA 91125. Phone: (626) 395-3785. Fax: (626) 440-9821. E-mail: avarsh@caltech.edu.

† Present address: Institute for Genetics, University of Cologne, Cologne D-50931, Germany.

‡ Present address: Department of Pediatrics and Psychiatry, University of Colorado School of Medicine, Denver, CO 80262.

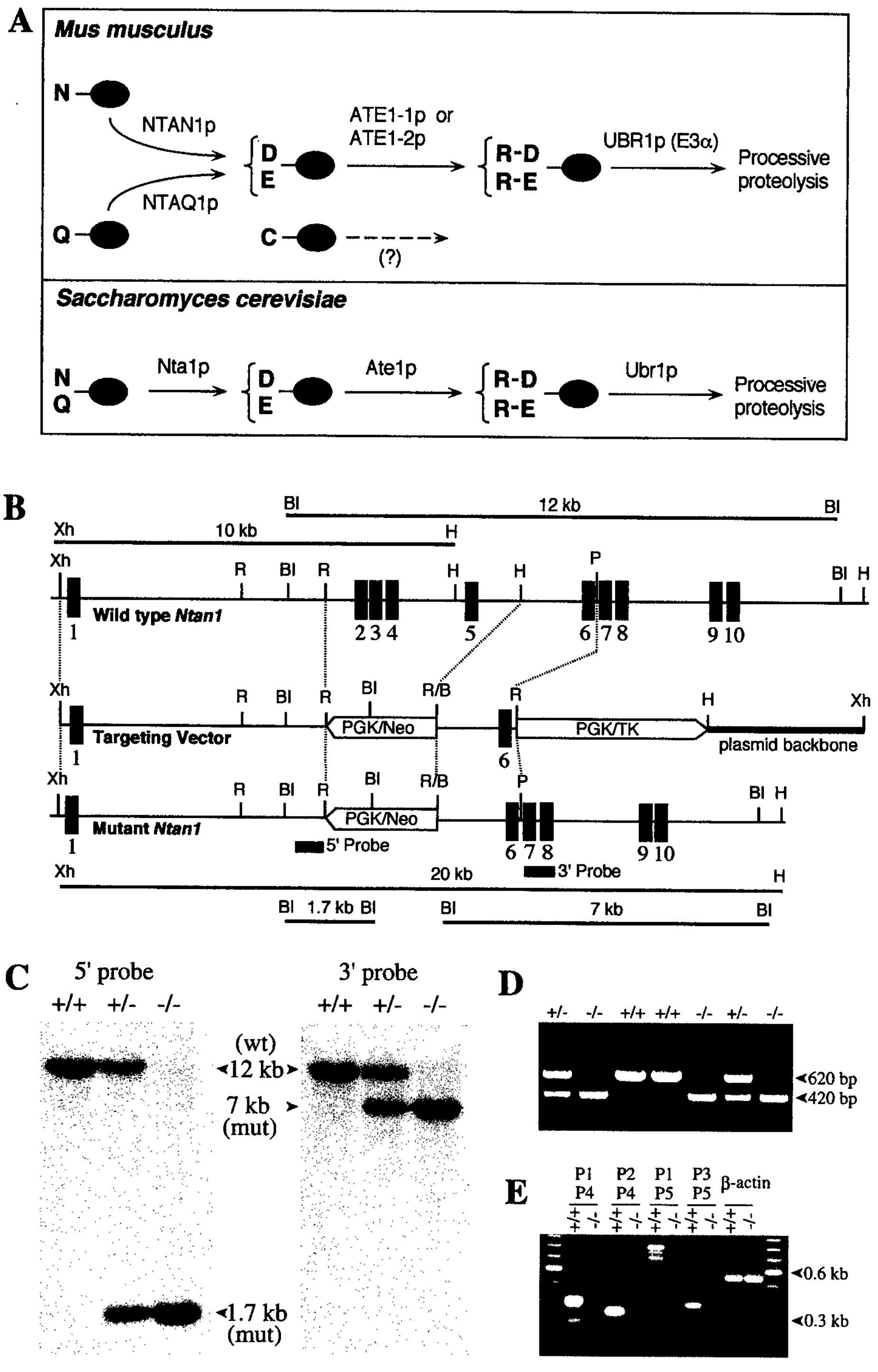

A
Mus musculus
N
NTAN1p
NTAQ1p
Q
D
E
ATE1-1p or
ATE1-2p
R-D
R-E
UBR1p (E3α)
Processive proteolysis
C
(?)
Saccharomyces cerevisiae
N
Q
Nta1p
D
E
Ate1p
R-D
R-E
Ubr1p
Processive proteolysis
B
BI 12 kb BI
Xh 10 kb H
Xh R BI R H H P BI H
Wild type Ntan1
1 2 3 4 5 6 7 8 9 10
Xh R BI R BI R/B R H Xh
PGK/Neo
PGK/TK
Targeting Vector
plasmid backbone
1 6
Xh R BI R BI R/B P BI H
PGK/Neo
Mutant Ntan1
5' Probe
1 6 7 8 9 10
3' Probe
Xh 20 kb H
BI 1.7 kb BI
BI 7 kb BI
C
5' probe
+/+ +/- -/-
3' probe
+/+ +/- -/-
(wt)
12 kb
7 kb
(mut)
1.7 kb
(mut)
D
+/- -/- +/+ +/+ -/- +/- -/-
620 bp
420 bp
E
P1 P4
P2 P4
P1 P5
P3 P5
β-actin
+/+ -/- +/+ -/- +/+ -/- +/+ -/- +/+ -/-
0.6 kb
0.3 kb

FIG. 1. Deletion-disruption of the mouse *Ntan1* gene. (A) Comparison of enzymatic reactions that underlie the activity of tertiary and secondary destabilizing residues in the yeast *S. cerevisiae* and the mouse. N-terminal residues are indicated by single-letter abbreviations for amino acids. The ovals denote the rest of a protein substrate. The *Ntan1*-encoded mammalian $\mathrm{Nt^{N}}$-amidase converts N-terminal Asn to Asp. N-terminal Gln is deamidated by $\mathrm{Nt^{Q}}$-amidase, which remains to be isolated (see text). In contrast, the yeast Nt-amidase Nta1p can deamidate either N-terminal Asn or Gln (6). The secondary destabilizing residues Asp and Glu are arginylated by the mammalian ATE1-1p or ATE1-2p R-transferase (34). A Cys-specific mammalian R-transferase (23) remains to be identified. N-terminal Arg, one of the primary destabilizing residues, is recognized by N-recognin, the E3 component of the N-end rule pathway (63). (B) Targeting strategy. Top, partial restriction map of the mouse *Ntan1* gene; middle, structure of the targeting vector; bottom, structure of the deletion-disruption $Ntan1^{-}$ allele. Exons are denoted by solid vertical bars. The directions of transcription of the *neo* and *tk* genes are indicated. Homologous recombination resulted in the replacement of the *Ntan1* exons 2 to 5 with the *neo* cassette. Probes for Southern hybridization are indicated by solid rectangles. Restriction sites: Xh, *Xho*I; R, *Eco*RI; BI, *Bam*HI; H, *Hin*dIII; P, *Pst*I. (C) Southern analysis of *Bam*HI-digested tail DNA from wild-type (+/+), heterozygous ($Ntan1^{+/-}$), and $Ntan1^{-/-}$ mice. The 5′ probe yielded the 12- and 1.7-kb *Ntan1* fragments for the wild-type (wt) and mutant (mut) *Ntan1* alleles, respectively; the 3′ probe detected 12- and 7-kb fragments. The organization of the deletion-disruption allele was independently verified by Southern analysis of the *Xho*I-*Hin*dIII-digested tail DNA (data not shown). (D) PCR analysis of tail DNA. The primers were 5′-GCCACTTGTGTAGCGCCAAGTGCCAGC (for *neo*, forward), 5′-CTTCCCACCAAGCCTGACTGTTGATC (for *Ntan1*, forward) and 5′-CTTCAATTTCTGTGCTCAGCTAAGCTC (for *Ntan1*, reverse). (E) RT-PCR analysis of the total RNA isolated from +/+ and $Ntan1^{-/-}$ EF cells, using primers P1 (for exon 1), P2 (exon 2), P3 (exon 6), P4 (exon 5), and P5 (exon 10). β-Actin mRNA was used as a control, at the 20-fold-lower primer concentration in comparison to other lanes.

strate-binding sites. The type 1 site binds the basic N-terminal residues Arg, Lys, and His. The type 2 site binds the bulky hydrophobic N-terminal terminal residues Phe, Leu, Trp, Tyr, and Ile (35, 63). Ubr1p contains yet another substrate-binding site that targets proteins such as Cup9p and Gpa1p, which bear internal (non-N-terminal) degrons (12, 54). The *Ubr1* genes encoding mouse and human N-recognins, also called E3α, have been cloned (36), and mouse strains lacking *Ubr1* have recently been constructed (Y. T. Kwon and A. Varshavsky, unpublished data).

The known functions of the N-end rule pathway include the control of peptide import in *S. cerevisiae*, through the degradation of Cup9p, a transcriptional repressor of *PTR2*, which encodes the peptide transporter (1, 12); a mechanistically undefined role in regulating the Sln1p-dependent phosphorylation cascade that mediates osmoregulation in *S. cerevisiae* (47); the degradation of alphaviral RNA polymerases and other viral proteins in infected metazoan cells (19, 38); and the degradation of Gpa1p, a Gα protein of *S. cerevisiae* (43, 54). Physiological N-end rule substrates were also identified among the proteins secreted into the mammalian cell's cytosol by intracellular parasites such as the bacterium *Listeria monocytogenes*. Short half-lives of these bacterial proteins are required for the efficient presentation of their peptides to the immune system (56). Inhibition of the N-end rule pathway was reported to interfere with mammalian cell differentiation (28) and to delay limb regeneration in amphibians (61). Studies of the Ub-dependent proteolysis of endogenous proteins in muscle extracts suggested that the N-end rule pathway plays a role in catabolic states that result in muscle atrophy (39, 57). A crush injury to the rat sciatic nerve was reported to result in a ~10-fold increase in the rate of arginine conjugation to the N termini of proteins in the nerve's region upstream of the crush site, suggesting an injury-induced increase in the concentration of R-transferase substrates and/or an enhanced activity of the N-end rule pathway (65).

Physiological substrates of either yeast or metazoan Nt-amidases and R-transferases are unknown. Engineered N-end rule substrates, including substrates of Nt-amidases and R-transferases, can be produced in vivo through the Ub fusion technique, in which a Ub-X reporter fusion is cleaved by deubiquitylating enzymes (DUBs) (66) after the last residue of Ub, yielding a reporter bearing the desired N-terminal residue X (3, 63).

The mouse Asn-specific Nt^{N}-amidase is encoded by the 17-kb *Ntan1* gene. The 1.6-kb *Ntan1* mRNA specifies the 310-residue Nt^{N}-amidase (24). In the present work, we characterized the expression and intracellular localization of Nt^{N}-amidase. We also constructed mouse strains bearing a homozygous deletion-disruption of *Ntan1* and showed that these mice lacked both Nt^{N}-amidase and the Asn-specific branch of the N-end rule pathway. The *Ntan1*$^{-/-}$ mice were fertile and outwardly normal but were found to differ from their congenic wild-type counterparts in spontaneous activity and spatial memory. Among these differences was a socially conditioned exploratory phenotype of *Ntan1*$^{-/-}$ mice that has not been previously described, to our knowledge, with other mouse strains.

MATERIALS AND METHODS

Construction of mouse strains lacking Nt^{N}-amidase. Genomic *Ntan1* DNA fragments were subcloned from the P1 phage DNA containing the strain C129-derived mouse *Ntan1* gene (24). The 1,462-bp *Hin*dIII-*Pst*I fragment containing exon 6 and its flanking introns was used as the vector's short homology arm. The 7.2-kb *Xho*I-*Not*I fragment containing exon 1 and its flanking intron was used as the long homology arm. To construct the targeting vector, a 1,462-bp *Hin*dIII-*Pst*I fragment containing exon 6 and its flanking introns (the vector's short arm) and the phosphoglycerate kinase gene (*PGK*)-thymidine kinase (*tk*) cassette of pPNT (a gift from R. C. Mulligan, Harvard Medical School, Boston, Mass.) were inserted into pPGKRN containing the wild-type *PGK-neo* (neomycin resistance gene) cassette (a gift from R. Jaenisch, Whitehead Institute, Cambridge, Mass.), yielding pPGK-SA. The vector's 7.2-kb long arm was produced by joining, within pPGK-SA, a 4.8-kb *Xho*I-*Eco*RI fragment to its flanking 2.6-kb *Eco*RI-*Eco*RI fragment containing exon 1 and flanking intron pPGK-SA, yielding the *Ntan1* targeting vector pNTAN1-KO (Fig. 1B). The *Xho*I-linearized targeting vector (Fig. 1B) was electroporated into CJ7 embryonic stem (ES) cells. Selection with G418 (at 0.4 mg/ml) and 1-(2′-deoxy, 2′ fluoro-β-D-arabinofuranosyl)-5-iodouracil (FIAU; at 0.4 μM) was started 24 h after electroporation. Correctly targeted ES cells were identified by PCR and Southern hybridization with the 5′ and 3′ probes (Fig. 1C and D). Cells of the 12 independent ES cell clones were injected into 3.5-days-postcoitum C57BL/6J blastocysts. The resulting male chimeras were bred with C57BL/6J females to test for germ line transmission of the mutated *Ntan1* gene. The *Ntan1*$^{+/-}$ mice resulting from this cross (6 out of 12 independent ES clones were found to populate germ line in these tests) were intercrossed to produce *Ntan1*$^{-/-}$ mice. Alternatively, the initial male chimeras were mated with 129/SvEv females, yielding, through the analogous series of steps, *Ntan1*$^{-/-}$ mice in the strain 129 background. All of the behavioral tests were carried out with *Ntan1*$^{-/-}$ mice in the strain 129 background. For genotyping, the tail-derived DNA was analyzed by PCR, or digested with either *Bam*HI or *Hin*dIII/*Xho*I, and analyzed by Southern hybridization. A 0.75-kb *Pst*I fragment and a 0.9-kb *Pst*I fragment (indicated in Fig. 1B) were used as the 5′ and 3′ hybridization probes, respectively.

Northern and RT-PCR analyses of RNA. Total RNA was isolated from brain, testis, and embryonic fibroblasts (EF cells) of wild-type and *Ntan1*$^{-/-}$ mice as described elsewhere (34). RNA was fractionated by electrophoresis in 1% formaldehyde-agarose gels, blotted onto Hybond N^{+} (Amersham), and hybridized with ^{32}P-labeled probes specific for different regions of the *Ntan1* cDNA (GenBank accession no. U57692): probe a, nucleotide (nt) 34 to 900; probe b, nt 118 to 450; probe c, nt 118 to 900; probe d, nt 34 to 450; probe e, nt 470 to 670; and probe f, nt 680 to 900. Alternatively, reverse transcription-PCR (RT-PCR) was carried out (2), with first-strand cDNA synthesized using Superscript II polymerase (GIBCO, Frederick, Md.). PCR was done using primers specific for different regions of the *Ntan1* cDNA: primer P1, 5′-ATGCCACTGCTGGTGGATGG-GCAG (forward); P2, 5′-GAGCCAGACTTCTCAGAGGTCAG (forward); P3, 5′-GACA-TTCACTTAGTGACATTATG (forward); P4, 5′-TTCTGTGACAGCTGCCTGTCATC (reverse); and P5, 5′-CATCAAGGTAGATCTAATATGTTC (reverse). The ratio of mouse *Ate1-1* and *Ate1-2* mRNAs was determined as described elsewhere (34).

Whole-mount in situ hybridization. Wild-type and *Ntan1*$^{-/-}$ embryos were staged, fixed, and processed for in situ hybridization as described elsewhere (16). A 0.3-kb fragment of the *Ntan1* cDNA (nt 108 to 448) that was encompassed by the deleted region in the *Ntan1*$^{-/-}$ allele was subcloned into *Xba*I-*Xho*I sites of pBluescript II SK^{+} (Stratagene), and the resulting plasmid, pMR27, was used as a template for synthesizing antisense RNA probe labeled with digoxigenin (Roche Molecular Biochemicals, Indianapolis, Ind.).

Localization of NTAN1p-GFP. The mouse *Ntan1* open reading frame (ORF) was subcloned into the *Xho*I and *Age*I sites of pEGFP-N1 (Clontech), yielding pNTAN1-GFP, which expressed the NTAN1p green fluorescent protein (GFP) fusion from the cytomegalovirus promoter. NIH 3T3 cells (ATCC 1658-CRL) were grown as monolayers in Dulbecco's modified Eagle medium (DMEM; Gibco-BRL) supplemented with 10% fetal bovine serum. Cells were grown to ~15% confluence on glass coverslips for 24 h prior to transfection with either the GFP-expressing control vector pEGFP-N1 or pNTAN1-GFP, using Lipofectamine (GIBCO) and the manufacturer-supplied protocol. Cells were incubated for 5 h at 37°C in serum-free DMEM containing DNA and Lipofectamine. An equal volume of medium containing 20% serum was then added, and the cells were grown for another 12 to 20 h at 37°C. Cells were fixed with 2% formaldehyde in phosphate-buffered saline, and GFP fluorescence was examined in a Zeiss Axiophot microscope.

Immortalization of EF cells, transfection, and pulse-chase analysis. EF cells were isolated from either *Ntan1*$^{-/-}$ or +/+ 13.5-day-old (e13.5) embryos as described elsewhere (52) and grown in DMEM–F-12 (GIBCO) supplemented with 15% fetal bovine serum, antibiotics, and 2 mM L-glutamine. They were immortalized using simian virus 40 (14) and then transiently transfected, using Lipofectamine PLUS (GIBCO), with plasmid pRC/dhaUbXnsP4βgal, which expressed DHFR (dihydrofolate reductase)-HA (hemagglutinin)-Ub^{R48}-X-nsP4βgal test proteins (X = Met, Asn, Gln, or Arg) (40) (see Results). Cells were labeled with Tran^{35}S-label (New England Nuclear, Boston, Mass.), followed by a chase for 0, 1, and 2 h in the presence of cycloheximide, preparation of extracts, immunoprecipitation, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), autoradiography, and quantitation with PhosphorImager, essentially as described elsewhere (40).

X-DHFR test proteins. ^{35}S-labeled X-DHFR proteins (X = Asn, Gln, or Asp) were prepared as described elsewhere (24). Briefly, plasmids pSG4, pSG41, and pSG44, expressing, respectively, Ub-Asn-DHFR, Ub-Asp-DHFR, and Ub-Gln-DHFR from the P_{trc} promoter (24), were transformed into *Escherichia coli* JM101 carrying pJT184, which expressed Ubp1p, a DUB of *S. cerevisiae* (62). Cells of a 50-ml culture grown at 37°C to an A_{600} of ~0.9 in Luria broth plus ampicillin (40 μg/ml) and chloramphenicol (20 μg/ml) were pelleted and resuspended in 50 ml of M9 supplemented with glucose (0.2%), thiamine (2 μg/ml), ampicillin (40 μg/ml), 1 mM isopropylthio-β-D-galactoside (IPTG), and methionine assay medium (Difco). The suspension was shaken for 1 h at 37°C, followed by the addition of 1 mCi of Tran^{35}S-label (ICN) and further incubation for 1 h at 37°C. Unlabeled L-methionine was then added to 1 mM, and shaking was continued for another 10 min. The cells were harvested, and lysates were prepared by the addition of lysozyme and Triton X-100, followed by centrifugation (24). The supernatant, containing [^{35}S]X-DHFR (X = Asn, Asp, or Glu) was purified by affinity chromatography on a methotrexate column (Pierce) (0.5-ml bed volume). [^{35}S]X-DHFRs were examined by SDS-PAGE and found to be greater than 95% pure.

IEF assay for amidase activity. Extracts from wild-type or $Ntan1^{-/-}$ EF cells were prepared by homogenization in a mixture containing 0.01% Triton X-100, 10% glycerol, 125 mM KCl, 7.5 mM MgCl$_2$, 1 mM dithiothreitol, 0.25 mM EDTA, 0.2 mM phenylmethylsulfonyl fluoride, 130 mM Tris-HCl (pH 7.5), and the protease inhibitors antipain, chymostatin, leupeptin, pepstatin, and aprotinin (Sigma), each at 25 μg/ml. For the deamidation assay, 5 μl of ^{35}S-labeled X-DHFR (0.5 mg/ml in storage buffer; X = Asn, Asp, or Glu) was mixed with 20 μl of the extract, incubated for 2 h at 37°C, and placed on ice. Samples (5 μl) were applied onto isoelectrofocusing (IEF)-PAGE plates (pH 3.5 to 9.5; Pharmacia, Piscataway, N.J.) precooled to 10°C. IEF was carried out for 80 min at 30 W in a cooled IEF apparatus (Hoefer, San Francisco, Calif.). The plates were soaked in 100 ml of 10% CCl$_3$COOH–5% 5-sulfosalicylic acid, stained with Coomassie blue to detect IEF markers (Pharmacia), and autoradiographed. ^{35}S in the bands of more acidic (deamidated) and more basic (initial) X-DHFR species was quantitated with a PhosphorImager (Molecular Dynamics, Sunnyvale, Calif.).

Behavioral tests. Mice were kept on a 0600- to 1800-h light cycle and tested during the light period. For the rotarod, weight retention, coat hanger, and platform-leaving tests, strain 129 $Ntan1^{-/-}$ mice and their congenic +/+ littermates (produced through matings of $Ntan1^{-/+}$ mice) were used. Nonlittermates were also used for the platform-leaving test (see below). For the shuttlebox and passive avoidance tests, the elevated plus-maze test, the open-field test, the Morris maze, the radial arm maze, and the Lashley maze, strain 129 +/+ and congenic $Ntan1^{-/-}$ mice (produced, respectively, through +/+ × +/+ and $Ntan1^{-/-}$ × $Ntan1^{-/-}$ matings) were used. The same series of tests were repeated 6 weeks or more after the initial experiment, using both male and female mice. To determine whether the *Ntan1* genotype affected long-term memory, mice were retested in the Morris maze, Lashley maze, and shuttlebox avoidance test at least 7 weeks after their original learning. The data were evaluated by analysis of variance. Significance was set at the 0.05 level; all behavioral differences described in Results were at this or a higher level of statistical significance.

(i) Rotarod test. The apparatus (UGO, Basile, Italy) consisted of a motor-driven rod 3 cm in diameter that carried five compartments, separated by walls 25 cm in height. One pair of littermates was placed on the rotating rod in each compartment. Mice were placed on the rod at either 10 or 20 rpm; the cutoff time was 2 min. Each mouse was tested on the rod six times, at 1-h intervals.

(ii) Weight retention. Mice were suspended by the tails and made to grab a wire loop (wire diameter, 2 mm; weight, 40 g) 70 cm above the floor. Time elapsed before the animal released the loop was measured. Each mouse was tested five times a day, at 1-h intervals, for a total of 3 days.

(iii) Coat hanger test. A wire coat hanger (wire diameter, 2 mm; length of side bars, 29 cm; length of the horizontal bar, 45 cm) suspended 75 cm above the table was used to examine motor coordination (11). A trial began by allowing the mouse to grab the middle of the horizontal bar with two front paws. The time elapsed before the animal grabbed the horizontal bar with all four paws and the time before the two front paws reached one of the side bars were determined. Each mouse in the trial was tested twice a day, at 6-h intervals, for 9 days.

(iv) Shuttlebox avoidance, passive avoidance, and elevated plus-maze tests. For the shuttlebox and passive avoidance tests, the Gemini avoidance system (San Diego Instruments, San Diego, Calif.) was used. The procedures are described in Results.

(v) Open-field test. For the open-field activity measurements, an Omnitech Digiscan animal activity monitor was used. Each mouse was placed into a square chamber (20 by 20 cm) for one 9-min session, and the movements were tabulated along the *x* and *y* axes, using infrared sensors.

(vi) Morris maze. A mouse was placed into a circular black water tub (diameter, 123 cm; temperature, 19 to 21°C) from one of the four quadrants and had to find a platform (diameter, 23 cm) submerged 1 cm below the surface of water and set 23 cm from the wall. There were no cues within the maze, and so the animal had to use extramaze spatial cues in guiding itself to the platform in repeated trials (46). An animal was given four trials per day (maximum of 45 s per trial), one from each of the four locations, their order determined quasi-randomly. Each mouse was tested for 5 days, with the platform in a fixed position. On day 6, the platform was moved to the quadrant diagonally opposite its previous location, and each mouse was given four trials, as before. This regimen is referred to as reversal learning. The mice were retested (with the platform in the original position) 7 or 8 weeks later. The animal's path was traced on an electronic digitizing tablet containing a template of the maze (22). The following parameters were recorded: total time, distance traveled, average speed, and percentage of time spent in each quadrant of the maze.

(vii) Spatial radial arm maze. The apparatus and testing procedure were described previously (30). Hidden escape platforms were placed at the ends of four of the eight identical arms which radiated from the central area. During training the animal was released from a start arm and had to find one of the escape platforms. Once a platform was found, the arm containing that platform was blocked. The training session consisted of four trials in which a mouse had to locate all four platforms. This was followed by 11 testing sessions which differed from the training session only by removing the platform while the mouse was in its home cage between trials. Thus, on successive trials, the animal had to remember which alleys it had previously entered and not enter those alleys again. The score was total errors obtained by summing together Working Memory Correct errors, Working Memory Incorrect errors, and Reference Memory errors (30).

(ix) Nonspatial radial arm maze learning. The nonspatial swimming radial arm maze apparatus and testing procedure were similar to the above spatial version, except for the following. Each of the arms was painted in a different black/white pattern: solid black, solid white, vertical stripes, horizontal stripes, black with white polka dots, white with black polka dots, zigzags, and the galvanized steel gray color (the start arm). The patterned arms which contained platforms were different for each mouse but remained constant throughout the testing of a given mouse. For the training session (session 1), the entrance to the arm with the just chosen platform was blocked; in addition, the maze was quasi-randomly rotated, so that each patterned arm now pointed toward a different place in space. As in the spatial version, the test sessions (sessions 2 to 12) were the same as the training session, except that while the mouse was in its home cage, each platform it found was removed from the maze, and the arm's entrance remained open for the remainder of the session. Over eight trials, or two sessions, each patterned arm pointed toward each of the eight spatial locations in the testing room. The errors were scored as above.

(x) Lashley maze learning. This maze contained cul-de-sacs that an animal had to learn to avoid, along with choices that required the animal to learn making correct left or right turns (T-choices). A water version of the maze was used, with the temperature at 19 to 20°C (21). All mice were given one trial per day for 5 days. An animal can learn this maze using two mutually nonexclusive strategies: by memorizing extramaze spatial cues, or by memorizing the sequence of correct left and right turns. The measures of learning included the learning index, defined as the number of correct entries divided by total number of entries, the number of cul entries (entries into dead ends) when swimming forward (i.e., toward the goal), number of forward T-choice errors, and the total number of backward errors. Retention testing was done 7 weeks later for experiment 1 and 8 weeks later for experiment 2.

(xi) Platform-leaving test. To compare the exploratory activities of two previously untested mice on the same platform, we devised the platform-leaving test. One $Ntan1^{-/-}$ mouse of strain 129 and one congenic +/+ mouse of the same gender (either a littermate or a nonlittermate) were placed, at the same time, on the platform (16 by 22 cm) 2.7 cm above the surface of a laboratory bench. For each test, the platform was covered with fresh white paper. Mice were allowed to either explore or step down from the platform up until the cutoff time of 3 min. (Step-down was recorded when all four paws of a mouse were no longer in contact with platform.) The number and genotypes of mice leaving the platform first, and the number of mice not leaving by 3 min, were recorded. A total of three trials, at 1-h intervals, were carried out for every pair of mice tested on a given day. Each set of tests employed ~10 pairs of mice 2 to 4 months old. Four independent sets of tests, with previously untested mice, were carried out over a period of 6 months.

RESULTS

Construction of $Ntan1^{-/-}$ mice. A deletion-disruption allele of the ~17-kb mouse *Ntan1* gene was constructed by replacing *Ntan1* exons 2 to 5 with a *neo* cassette (Fig. 1B). Exons 2 to 5 encoded a 118-residue region of the 310-residue NTAN1p (Nt^{N}-amidase) that was significantly conserved among the NTAN1p proteins of different species, including the plant *Arabidopsis thaliana*, the fly *Drosophila melanogaster*, and mammals (data not shown). Of the ~1,000 ES cell clones (strain 129/CJ7) resistant to both G418 and FIAU, 33 clones contained the expected mutation, as verified by PCR and Southern analyses (data not shown). Twelve of these correctly targeted ES cell clones were used to generate male chimeras, and in six of them the $Ntan1^{-}$ allele was transmitted through the germ line. Male chimeras were mated with either 129/SvEv or C57BL/6 females, yielding $Ntan1^{+/-}$ heterozygotes. Intercrosses of heterozygous mice produced $Ntan1^{-/-}$ progeny at the expected frequency of approximately one in four (Fig. 1C and D), indicating that the absence of NTAN1p did not increase embryonic lethality. The behavioral tests described be-

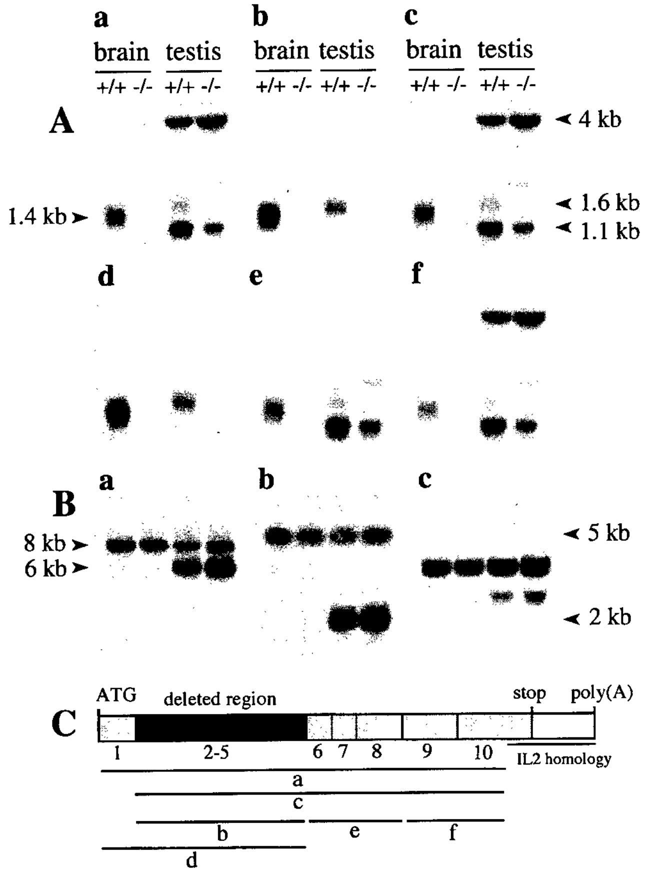


FIG. 2. Northern hybridization analysis of the total RNA isolated from +/+ and *Ntan1*$^{-/-}$ brains and testes. (A) Hybridization using probes a to f that encompassed different regions of the *Ntan1* cDNA (indicated in panel C). (B) The same Northern blots were hybridized with probes specific, respectively, for the mouse *Ubr1*, *Ate1*, and β-actin cDNAs. (C) Exons of *Ntan1* and the hybridization probes used. The sequence of a 208-bp segment of the *Ntan1* cDNA (nt 896 to 930), termed the IL-2 homology region, is 98.6% identical to the sequence of a 206-bp segment in the 3′-flanking untranslated region of the mouse *Il2* gene, which encodes IL-2 (24).

low used exclusively *Ntan1*$^{-/-}$ mice produced through matings in the strain 129 background.

Expression of *Ntan1* and intracellular localization of NTAN1p. Several regions of *Ntan1* transcripts were analyzed using RT-PCR with RNA from wild-type and *Ntan1*$^{-/-}$ EF cells. The level of *Ntan1*-derived RNA in the *Ntan1*$^{-/-}$ EF cells was below the RT-PCR detection threshold (Fig. 1E and data not shown). Northern hybridization analyses, using probes encompassing either exons 2 to 5 (the deleted region) or exons 1 to 5 detected the 1.4- and 1.6-kb *Ntan1* mRNAs, respectively, in the brains and testes of +/+ mice; the +/+ testes also contained the *Ntan1*-derived RNAs of 1.1 and 4 kb (Fig. 2A). No *Ntan1* transcripts were detected in the brains and testes of *Ntan1*$^{-/-}$ mice with these probes (Fig. 2Ab and d).

Similar analyses but using probes encompassing exons 1 to 10 or 2 to 10 (except for the interleukin-2 [IL-2] similarity region of exon 10 [24]) detected at most trace amounts of *Ntan1*-derived transcripts in the brains of *Ntan1*$^{-/-}$ mice (Fig. 2Aa and c). However, in the testes of *Ntan1*$^{-/-}$ mice, the same probes detected 1.1- and 4-kb *Ntan1*-derived, testis-specific transcripts but not the 1.6-kb transcript (Fig. 2Aa and c). Further mapping, using the probes shown in Fig. 2C, indicated that the 1.1- and 4-kb transcripts hybridized to *Ntan1* exons 6 to 10 and exons 9 and 10, respectively (Fig. 2Ae and f). These transcripts were detected with antisense but not sense probes (data not shown), indicating a direction of transcription identical to that for the full-length *Ntan1* mRNA. Thus, the testes but not the brains of *Ntan1*$^{-/-}$ mice contained a set of *Ntan1*-derived transcripts that lacked the 5′ half of the *Ntan1* ORF (including the region deleted in *Ntan1*$^{-/-}$ mice) and encompassed its 3′ half. Apparently the same transcripts were present in the +/+ testes (Fig. 2Ae and f). Whether these testis-specific transcripts are physiologically relevant remains to be determined.

We also examined by Northern analysis whether the expression of other components of the N-end rule pathway was altered in the *Ntan1*$^{-/-}$ mice. No changes in the expression of either *Ubr1* (encoding the E3 of the N-end rule pathway) or *Ate1* (encoding R-transferase [34]; see the introduction) were detected in the brains and testes of *Ntan1*$^{-/-}$ mice (Fig. 2B). The expression of *Ubr2* and *Ubr3* (homologs of *Ubr1*; Kwon and Varshavsky, unpublished data) and of *mHR6B*, encoding the $E2_{14K}$ Ub-conjugating (E2) enzyme, was also unchanged in the *Ntan1*$^{-/-}$ mice (data not shown). In addition, no alterations in the ratio of *Ate1-1* mRNA to *Ate1-2* mRNA, which encode two splicing-derived forms of the mouse R-transferase (34), was detected in the brains, testes, and EF cells of *Ntan1*$^{-/-}$ mice (data not shown).

Ntan1 was expressed in most if not all tissues of adult mice (24). Whole-mount in situ hybridization was used to examine the expression of *Ntan1* in +/+ mouse embryos. In e9.75 embryos, *Ntan1* was most strongly expressed in the branchial arches, the tail bud, and the forelimb buds; the same probe detected no signal in the *Ntan1*$^{-/-}$ embryos (Fig. 3A). In e10.5 and e11.5 embryos, the expression of *Ntan1* became high in the hindlimb buds as well; this expression pattern was indistinguishable from that of *Ubr1* (Fig. 3A), consistent with the assignment of NTAN1p and UBR1p to the same pathway.

To determine the intracellular location of NTAN1p, it was expressed in mouse 3T3 cells as a fusion to the N terminus of GFP. Whereas the free 26-kDa GFP was distributed uniformly throughout the cell, the 65-kDa NTAN1p-GFP was enriched in the nucleus, was also present in the nucleus-proximal cytoplasm, but was apparently much less abundant at the cells' periphery (Fig. 3B and data not shown).

***Ntan1*$^{-/-}$ mice lack Nt^{N}-amidase and the asparagine-specific branch of the N-end rule pathway.** To determine whether *Ntan1*$^{-/-}$ EF cells lacked Nt^{N}-amidase activity, we used a previously described assay (24). The ^{35}S-labeled X-DHFR test proteins (X = Asn, Gln, or Asp) were incubated with a whole-cell extract and thereafter fractionated by IEF, which separated Asn-DHFR and Gln-DHFR from Asp-DHFR and Glu-DHFR, respectively. Incubation of Asn-DHFR with the extract from +/+ EF cells shifted the pI of this protein to that of Asp-DHFR. In contrast, the pI of Asn-DHFR remained unchanged after an otherwise identical incubation with the same amount of extract from *Ntan1*$^{-/-}$ EF cells (Fig. 4A). The pI of Gln-DHFR shifted to that of Glu-DHFR after incubations with either +/+ or *Ntan1*$^{-/-}$ extracts (Fig. 4A). Thus, *Ntan1*$^{-/-}$ EF cells lacked Nt^{N}-amidase activity but contained wild-type amounts of the Gln-specific Nt^{Q}-amidase. The latter result indicated that the putative mammalian Nt^{Q}-amidase (which remains to be isolated and cloned) is encoded by a gene distinct from *Ntan1*.

To examine the in vivo degradation of N-end rule substrates in the absence of Nt^{N}-amidase, the immortalized +/+ and *Ntan1*$^{-/-}$ EF cell lines were transiently transfected with plasmids that expressed X-nsP4βgal test proteins (X = Met, Asn, Gln, or Arg). An X-nsP4βgal was expressed as part of a fusion containing the HA-tagged reference protein DHFR. In this

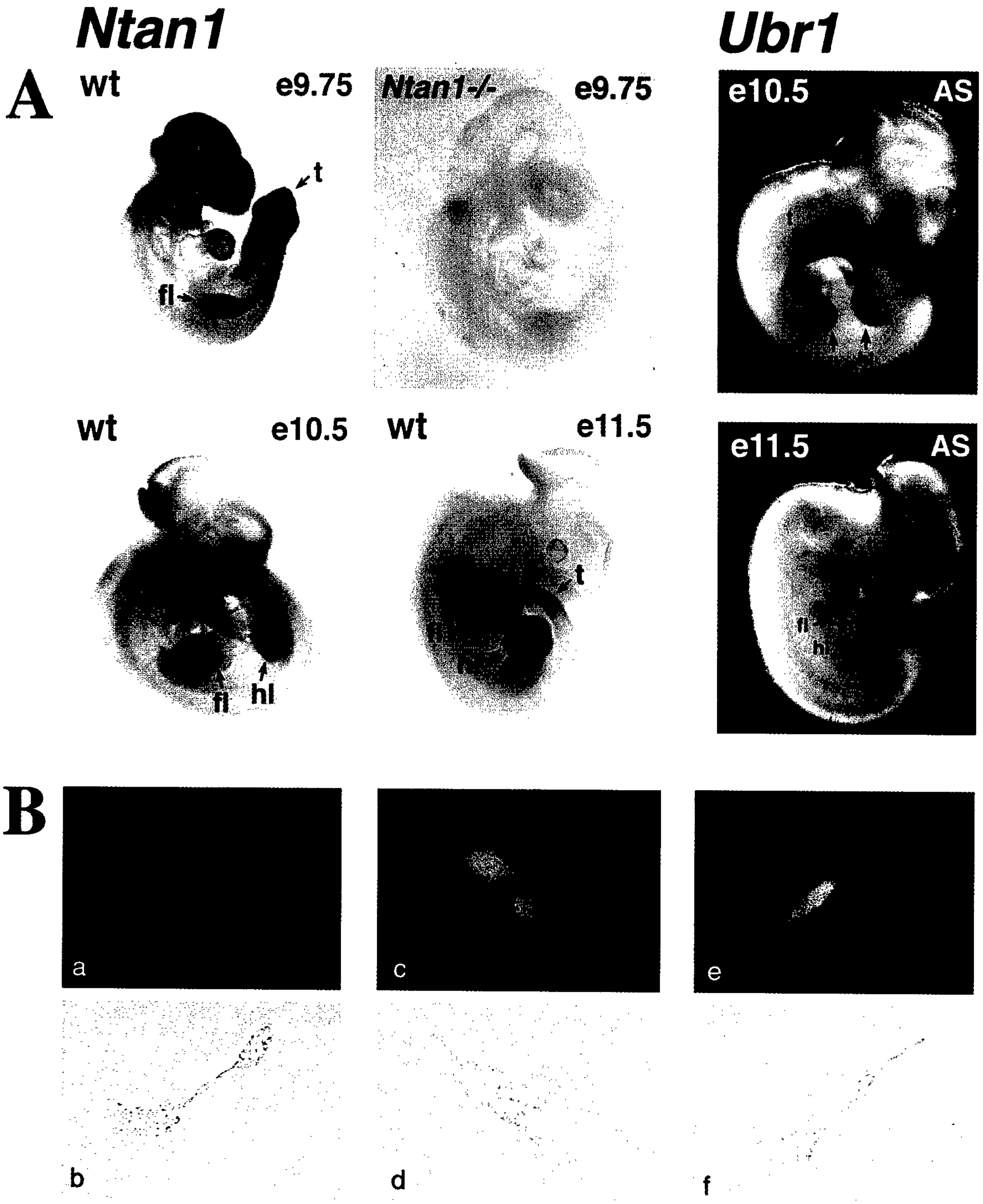


FIG. 3. Expression of *Ntan1* mRNA and localization of NTAN1p. (A) Whole-mount in situ hybridization of wild-type (wt) and *Ntan1*$^{-/-}$ embryos (left four panels, light background) with an antisense RNA probe derived from a 0.3-kb fragment of the *Ntan1* cDNA (nt 108 to 448) that was absent from the *Ntan1*$^{-/-}$ allele. The regions of high *Ntan1* expression in the tail buds (t), forelimb buds (fl), and hindlimb buds (hl) are indicated. The right two panels (dark background) show the results of in situ hybridization with an antisense (AS) *Ubr1* cDNA probe. (B) Intracellular localization of mouse NTAN1p. NIH 3T3 cells were transiently transfected with a plasmid expressing the NTAN1p-GFP fusion (see Materials and Methods). Typical fluorescence patterns (a, c, and e) and the matching phase-contrast images (b, d, and f) are shown (see text).

previously developed UPR (Ub protein reference) technique (40, 60), a DHFR-HA-UbR48-X-nsP4βgal fusion is cotranslationally cleaved by DUBs at the Ub-X-nsP4βgal junction, yielding the long-lived DHFR-HA-UbR48 reference protein and the test protein X-nsP4βgal. In the UPR technique, the reference protein serves as an internal control for the levels of expression, immunoprecipitation yields, sample volumes, and other sources of sample-to-sample variation, thereby increasing the accuracy of pulse-chase assays (40, 60).

The previously introduced term ID^x (initial decay, i.e., the extent of degradation of a protein during the pulse of *x* minutes [40]) was used to describe the in vivo decay curves of test

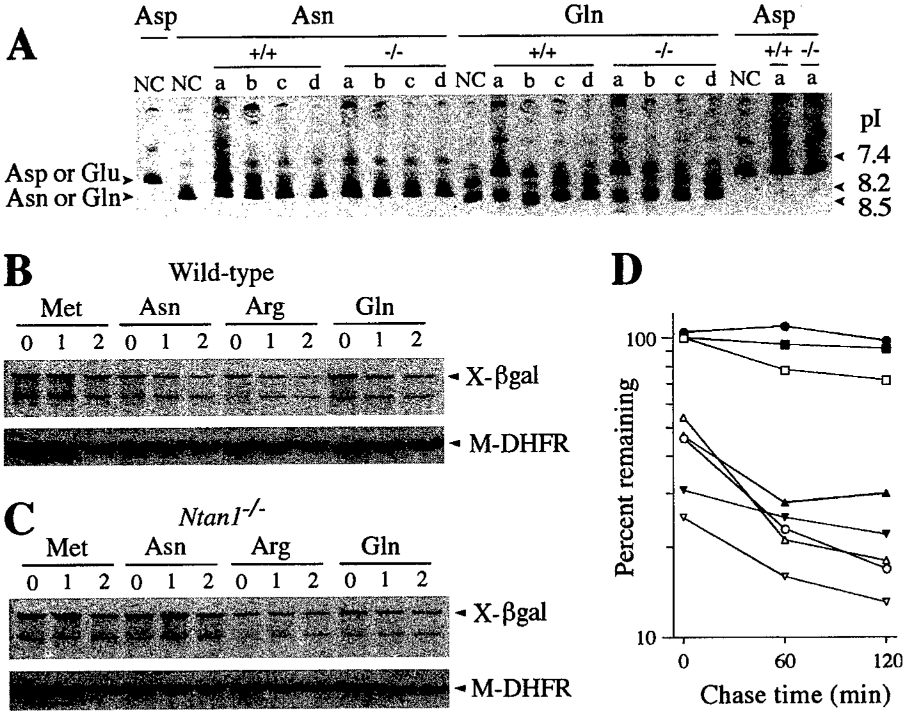


FIG. 4. Mouse *Ntan1*$^{-/-}$ cells lack Nt^{N}-amidase and are unable to degrade the normally short-lived N-end rule substrates bearing N-terminal Asn. (A) ^{35}S-labeled, purified X-DHFR test proteins (X = Asn, Gln, or Asp) were incubated for 2 h at 37°C with buffer alone (negative controls [NC]) or with extracts from either wild-type or *Ntan1*$^{-/-}$ EF cells, followed by IEF and autoradiography. The assays were carried using either the initial extracts (lanes a) or the same extracts diluted with buffer 10-, 100-, and 1,000-fold (lanes b, c, and d, respectively). The IEF positions of X-DHFRs bearing N-terminal Asp or Glu versus Asn or Gln are shown on the left. The corresponding pH values are indicated on the right. (B) Immortalized +/+ and *Ntan1*$^{-/-}$ EF cells (see Materials and Methods) were transiently transfected with plasmid pRC/dhaUbXnsP4βgal, which expressed DHFR-HA-Ub^{R48}-X-nsP4βgal test proteins (X = Met, Asn, Gln, or Arg) (40). These proteins were cotranslationally cleaved in vivo by DUBs, yielding the long-lived reference protein DHFR-HA-Ub^{R48} and the test protein X-nsP4βgal (X = Met, Asn, Gln, or Arg) (see text). Cells were labeled with [^{35}S]methionine-cysteine, followed by a chase for 0, 1, and 2 h (as indicated at the top) in the presence of cycloheximide, preparation of extracts, immunoprecipitation, SDS-PAGE, autoradiography, and quantitation, essentially as described elsewhere (40). The bands of X-nsP4βgal proteins and the DHFR-HA-Ub^{R48} reference protein are indicated on the right as X-βgal and DHFR. (C) Same as panel B but with immortalized *Ntan1*$^{-/-}$ EF cells. (D) Quantitation of the in vivo degradation of X-nsP4βgal test proteins using the reference-based pulse-chase patterns in panels B and C (see Materials and Methods). The amounts of ^{35}S in an X-nsP4βgal protein, relative to ^{35}S in the DHFR-HA-Ub^{R48} reference protein at the same time points, were plotted as percentages of this ratio for Met-nsP4βgal at time zero. (Met-nsP4βgal bore a stabilizing N-terminal residue.) Open and closed symbols, wild-type and *Ntan1*$^{-/-}$ EF cells, respectively; □ and ■, Met-nsP4βgal; ○ and ●, Asn-nsP4βgal; ▽ and ▼, Arg-nsP4βgal; △ and ▲, Gln-nsP4βgal.

proteins. The term ID^{x} is superfluous in the case of a strictly first-order decay, which is defined by a single half-life. However, the in vivo degradation of most proteins deviates from first-order kinetics. Specifically, the rate of degradation of short-lived proteins can be much higher during the pulse, in part because a newly labeled (either nascent or just completed) polypeptide is conformationally immature and consequently may be targeted for degradation more efficiently than its mature counterpart. This enhanced early degradation, previously termed the zero-point effect (5), is described by the parameter ID^{x} (40). It was found that a large fraction of the zero-point effect results from the cotranslational degradation of nascent (being synthesized) polypeptides, which never reach their mature size before their destruction by processive proteolysis (G. Turner and A. Varshavsky, unpublished data).

Asn-nsP4βgal was metabolically unstable in +/+ EF cells: its ID^{10} (percent degradation during the 10-min pulse) was ~45%, in comparison to ID^{10} of the long-lived Met-nsP4βgal, taken as 100% (Fig. 4B and D). However, the same Asn-nsP4βgal was completely (and reproducibly) stabilized in the *Ntan1*$^{-/-}$ EF cells, its pulse-chase pattern becoming indistinguishable from that of the normally long-lived Met-nsP4βgal. In contrast, both Gln-nsP4βgal and Arg-nsP4βgal were comparably short-lived in both +/+ and *Ntan1*$^{-/-}$ cells (Fig. 4B to D). Taken together with the measurements of Nt^{N}-amidase enzymatic activity (Fig. 4A), these data indicated that the *Ntan1*$^{-/-}$ EF cells, in contrast to congenic wild-type cells, lacked the Asn-specific branch of the N-end rule pathway but retained the rest of this pathway.

Motor coordination and spontaneous activity of *Ntan1*$^{-/-}$ mice. The *Ntan1*$^{-/-}$ mice were fertile, apparently healthy, and of similar size and weight as +/+ littermates, and they cared for their offspring. These mice oriented to sound; their limb movements and behavior appeared to be indistinguishable from those of congenic +/+ mice. Histological examination of *Ntan1*$^{-/-}$ tissues (small intestine, liver, pancreas, adrenal gland, thyroid gland, kidney, ovary, testis, heart, spleen, thymus, skeletal muscle, brain, and sciatic nerve) did not detect abnormalities. No significant differences were observed between +/+ and *Ntan1*$^{-/-}$ thymocytes in the ability to undergo apoptosis in response to either radiation or dexamethasone, suggesting that *Ntan1*$^{-/-}$ mice were not impaired in apoptotic responses. Thus far, the only nonbehavioral differences between +/+ and congenic *Ntan1*$^{-/-}$ mice were a weaker mitogenic response of *Ntan1*$^{-/-}$ splenocytes and thymocytes to phytohemagglutinin and hypersensitivity of *Ntan1*$^{-/-}$ mice to bacterial lipopolysaccharide (data not shown). In addition, over the last 1.5 years the frequency of natural death among *Ntan1*$^{-/-}$ mice was ~50% higher than among congenic +/+ mice.

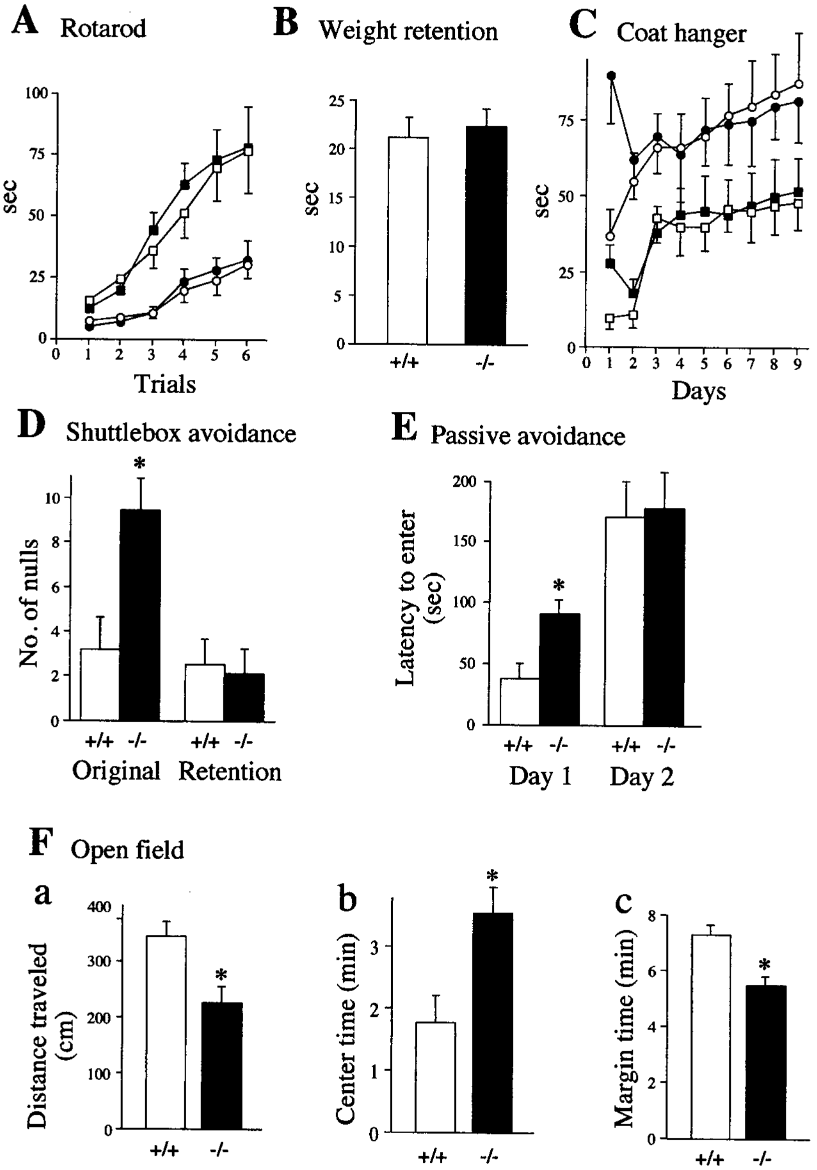


FIG. 5. Normal motor coordination and reduced spontaneous activity of *Ntan1*$^{-/-}$ mice. Open symbols: +/+ mice; solid symbols: congenic *Ntan1*$^{-/-}$ mice (see Materials and Methods). Statistically significant differences ($P < 0.05$) are indicated by *. Experiments used 24 *Ntan1*$^{-/-}$ mice and their congenic +/+ littermates (A to C) or nonlittermates (D to G). (A) Rotarod test. Data show time elapsed before the animals fell from a horizontal rod rotating at 10 rpm (squares) or 20 rpm (circles). No significant differences were found between +/+ and *Ntan1*$^{-/-}$ mice. (B) Weight retention test. Data show time elapsed before the animals released a hook. The height of each bar represents the time averaged from 360 trials. There were no significant differences between +/+ and *Ntan1*$^{-/-}$ mice. (C) Coat hanger test. Squares, time elapsed before the animal grabbed the horizontal bar with both rear paws; circles, time elapsed before the front two paws reached one of the side bars. Each data point is an average of 48 trials. There were no significant differences between +/+ and *Ntan1*$^{-/-}$ mice. (D) Shuttlebox avoidance test. Shown are the number of null responses per 50 daily trials in which the mouse remained in the original compartment of the shuttlebox and received 20 s of electric shock during the original learning and retention testing 7 to 10 weeks later. (E) Passive avoidance test, measuring time before entering a dark chamber after an automated guillotine door opened, exposing the dark chamber. On day 1 the mice were shocked after entering the dark chamber; 13 *Ntan1*$^{-/-}$ mice and 12 congenic +/+ mice were used. (F) Open-field test. Shown are the total distance traveled by the animal during the observation time of 9 min (a), time (out of 9 min in total) that the animal spent in the center area (b), and time (out of 9 min in total) that the animal spent within 1 cm of the walls. This set of tests used 28 *Ntan1*$^{-/-}$ mice and 27 congenic +/+ mice.

Motor coordination of *Ntan1*$^{-/-}$ and congenic +/+ mice was compared using the rotarod apparatus. No difference in the ability of mice to stay on a rotating horizontal rod was observed between the *Ntan1*$^{-/-}$ and +/+ strains (Fig. 5A). Comparisons of physical strength (weight retention test), of both coordination and strength (coat hanger test), and of walking patterns (hindpaw footprint test) also did not distinguish between *Ntan1*$^{-/-}$ and +/+ mice (Fig. 5B and C and data not shown). These and related observations indicated that *Ntan1*$^{-/-}$ mice were not impaired in either motor coordination, physical strength, or the learning ability required for improved performance in repeated trials (Fig. 5A and C).

The locomotor activity of *Ntan1*$^{-/-}$ mice was assayed in the open-field test. By several criteria, *Ntan1*$^{-/-}$ mice exhibited significantly lower spontaneous activity than the congenic wild-type mice (Fig. 5F). Specifically, *Ntan1*$^{-/-}$ mice traveled significantly less distance (Fig. 5Fa), spent significantly more time near the field's center (Fig. 5Fb), and consequently spent less time within 1 cm from the walls (Fig. 5Fc). These findings will be considered in conjunction with data from the platform-leaving test (see Fig. 6 and Discussion).

Active and passive avoidance learning. The shuttlebox avoidance test compared shock-motivated learning in +/+ and congenic *Ntan1*$^{-/-}$ mice. In this active avoidance learning procedure, a light-conditioned stimulus precedes, by 5 s, the unconditioned stimulus of electric shock, which remains on for 20 s. The animal can (i) cross to the safe compartment within the 5-s conditioned-stimulus interval, thus avoiding shock; (ii) cross over during the 20-s unconditioned-stimulus interval, to escape the ongoing shock; or (iii) not cross over and receive the full 20 s of shock (the latter outcome is scored as the null response). Neither +/+ nor *Ntan1*$^{-/-}$ mice showed any evidence of learning over the 5 days of testing. However, there were performance differences in the distribution of their responses. The two groups averaged only four avoidance responses per day, but the +/+ mice exhibited more escape responses and therefore fewer null responses than the *Ntan1*$^{-/-}$ mice (Fig. 5D). Neither group showed evidence of learning during the 5 days of retention testing, but the number of null responses by *Ntan1*$^{-/-}$ mice decreased significantly in comparison to their performance during learning (data not shown), indicating that memory consolidation had occurred in this group between the original testing and retesting 7 to 10 weeks later. Strain 129 mouse substrains are known to perform poorly in active avoidance tests (7, 53).

In the shuttlebox text, an electric shock is used to compel the animal to make an active response. By contrast, in passive learning, the shock is used to compel the animal to inhibit its behavior. The mouse was placed into a chamber. Ten seconds later, the chamber was illuminated and a guillotine door opened, exposing a dark chamber. After the animal moved into the dark (the natural behavior), the guillotine door closed and the animal received 0.4 mA of foot shock for 2 s. The time it took the mouse to enter the dark chamber (until the cutoff time of 400 s) was recorded. The mouse was put back in its home cage; 24 h later, the mouse was placed back into the original chamber, and all stimulus conditions were repeated except that no shock occurred when the mouse entered the dark chamber. The usual finding is that the mouse takes longer to enter the dark chamber on the second day because the prior shock experience inhibits its behavior. Figure 5E shows the time before entry into the dark chamber (latency time) on both days for the two genotypes. On the first day, the *Ntan1*$^{-/-}$ mice took significantly longer to enter the dark chamber than the +/+ animals. On the second day, the entry into the dark chamber by the *Ntan1*$^{-/-}$ mice was further delayed (indicating learning), while the entry by the +/+ mice was delayed even more, resulting in similar retention times for either genotype (Fig. 5E).

The longer retention time of the *Ntan1*$^{-/-}$ mice on the first day of passive learning, as well as their lower activity in the open-field test, suggested that +/+ and congenic *Ntan1*$^{-/-}$ mice might differ in their anxiety levels. To investigate this, they were tested in the elevated plus maze, which consists of two open arms and two closed arms radiating from a center platform. The open arms are more anxiety inducing, in that the animals with higher anxiety tend to decrease exploration of the open arms (42). Although the *Ntan1*$^{-/-}$ mice tended to explore the open arms less frequently than congenic +/+ mice, the observed difference was not statistically significant (data not shown).

A socially conditioned exploratory phenotype of *Ntan1*$^{-/-}$ mice. To address the finding of lower activity in the *Ntan1*$^{-/-}$ mice (Fig. 5F) in a different way, we devised a simple test in which the exploratory activity was assessed in the context of a social interaction. In this platform-leaving test, two previously untested mice, *Ntan1*$^{-/-}$ and congenic +/+, are placed, close together and at the same time, onto the center of a 16- by 22-cm platform 2.7 cm in height. Thereafter one records the following variables: the time it takes each mouse to leave the platform, until the cutoff time of 3 min; the number and genotype of mice leaving the platform first (in independent trials with pairs of mice); and the number and genotypes of mice not leaving the platform by the cutoff time (see Materials and Methods). In contrast to the standard open-field test, a mouse in the platform-leaving test sees and smells the second mouse on the same platform and is otherwise influenced by the second mouse in the course of their peregrinations over the platform. Thus, the proclivity of a mouse to explore its environment is modulated, in the platform-leaving test, by interactions with another mouse.

A striking result of this test was the observation that an *Ntan1*$^{-/-}$ mouse paired with its congenic +/+ littermate tended to stay longer on the platform and almost never left it by the cutoff time, in contrast to the +/+ mouse (Fig. 6A). For example, in a trial with 44 pairs of previously untested *Ntan1*$^{-/-}$ and +/+ mice, only 1 *Ntan1*$^{-/-}$ mouse left the platform by the cutoff time, whereas 20 +/+ mice left it by that time (Fig. 6Ac). Repeated trials, at 1-h intervals, in the platform-leaving test progressively decreased this initial difference between the paired *Ntan1*$^{-/-}$ and +/+ mice (Fig. 6A). Note that this decrease resulted from both an increase in the exploratory activity of *Ntan1*$^{-/-}$ mice (in successive trials) and a decrease in the exploratory activity of their paired +/+ littermates (e.g., Fig. 6Ac). Thus, the strongly different initial responses of paired *Ntan1*$^{-/-}$ and +/+ mice to the novel environment (of which the other mouse was a part) became attenuated in subsequent trials (Fig. 6A) and nearly disappeared by the seventh trial (data not shown).

We then asked whether the observed difference between *Ntan1*$^{-/-}$ and +/+ mice was retained if the same test was performed with congenic nonlittermates. Strikingly, if the test of Fig. 6A was carried out with wild-type and mutant nonlittermates, the results were reproducibly different from those with littermates (Fig. 6B). Specifically, in this case *Ntan1*$^{-/-}$ mice did not stay longer on the platform: they left it, on average, more frequently than +/+ mice, a pattern opposite that seen with littermates (Fig. 6Bb; compare with Fig. 6Ab). In addition, the striking difference between *Ntan1*$^{-/-}$ and +/+ littermates in their frequency of not leaving the platform by the cutoff time was no longer seen in the otherwise identical tests with nonlittermates (Fig. 6Ac; compare with Fig. 6Bc).

In the control experiments, multiple pairs of littermates of the same genotype (either both +/+ or both *Ntan1*$^{-/-}$) were subjected to this test. No statistically significant differences between the platform-leaving patterns of the two mice in a pair were observed (data not shown). The results presented in Fig. 6 were consistently reproduced in four independent sets of tests, carried out over 6 months with previously untested mice (at least 10 pairs of mice per test). This socially conditioned exploratory phenotype of *Ntan1*$^{-/-}$ mice (Fig. 6) has not been previously described, to our knowledge, with other mouse strains.

Spatial learning and memory. To compare spatial learning of +/+ and congenic *Ntan1*$^{-/-}$ mice, we tested them on the hidden platform Morris maze (where a mouse had to find a platform submerged in water pool), using the reversal learning regimen (see Materials and Methods) and retesting 7 or 8 weeks later. Both +/+ and *Ntan1*$^{-/-}$ mice learned to find the platform and did not differ in either the time taken to do it or

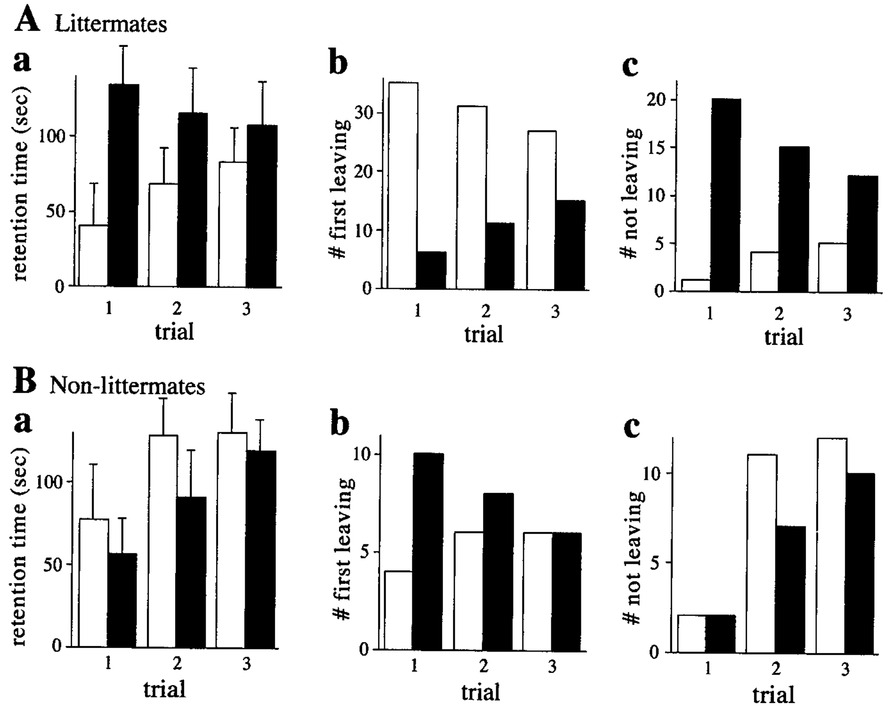


FIG. 6. Socially conditioned differences in exploratory behavior between +/+ and congenic *Ntan1*$^{-/-}$ mice. Open bars, +/+ mice; solid bars, congenic *Ntan1*$^{-/-}$ mice. In the platform-leaving test, one *Ntan1*$^{-/-}$ mouse and one +/+ mouse (either a littermate or a nonlittermate) were placed, at the same time, on a 16- by 22- by 2.7-cm platform. Mice were allowed to either explore or step down from the platform until the cutoff time of 180 s. Bars show the time by which each animal left the platform (a), the number of animals, of each genotype, leaving first (b), and the number of animals, of each genotype, not leaving by the cutoff time (c). Four independent sets of tests (at least 10 pairs of mice per test) were carried out over 6 months, using 44 pairs of +/+ and *Ntan1*$^{-/-}$ littermates, which were produced through heterozygous matings and identified by genotyping ~700 mice. The results differed by less than 15% among the independent tests. (B) Same as panel A except with pairs of nonlittermates (24 +/+ mice and 24 *Ntan1*$^{-/-}$ mice).

the distance traveled; they also did not differ on these measures in the subsequent reversal learning test, when the platform was moved to the diagonally opposite quadrant (data not shown). However, 7 to 8 weeks later, on the first day of retention testing, the *Ntan1*$^{-/-}$ mice were inferior to congenic +/+ mice, in that they took significantly longer time to find the platform and traveled a significantly longer distance before reaching it (data not shown), suggesting that the *Ntan1*$^{-/-}$ mice have a less effective spatial memory system.

Another measure of spatial learning employs the water version of the radial arm maze, where four of the eight arms contain escape platforms (30). There are no intramaze cues available to the mouse, and it must use extramaze spatial information involving reference and working memory to remember which alleys to enter and which to avoid. The mice were given training for 1 day and tested for 11 consecutive days, receiving four trials each day. The measure used was the total number of errors. *Ntan1*$^{-/-}$ mice made significantly more errors in the acquisition phase (sessions 2 to 7) than the congenic +/+ mice (Fig. 7A). In the asymptotic phase (sessions 8 to 12), there were no significant differences between the two strains (Fig. 7A).

The four trials within each daily session impose an increasing demand on memory. Thus, on the first trial, there are platforms in four of the eight alleys, and entering any of these terminates the trial. On the second trial, three of the alleys still contain platforms; the mouse has to remember which alley it had entered on the prior trial and not enter that alley again. Finally, on the fourth trial, the mouse has to remember which three alleys it had chosen earlier, so that it can find the one remaining alley containing a platform. Thus, by studying the error rate from trial to trial, summed over the sessions, one can compare the abilities of mice to cope with an increasing memory load. Figure 7C shows the error curves summed over the acquisition phase. As expected, errors increased as the memory load increased. However, the rate of increase was significantly greater for the *Ntan1*$^{-/-}$ mice, especially on trial 4 (Fig. 7C).

Unlike previously tested mouse strains (30), both +/+ and congenic *Ntan1*$^{-/-}$ mice exhibited poor learning over the 11 days of testing, consistent with inferior performance of strain 129-derived mice in several other learning regimens (7, 53). Thus, the significant effects described above (Fig. 7A and C) represent largely differences in performance but indicate little about the spatial learning ability of these mice. To determine whether the observed effects were related to spatial competence, the identical experiment was carried out using the same radial arm maze, except that a nonspatial version was devised, by having different visual patterns in each alley and rotating the maze between trials so that the extramaze cues were rendered meaningless. In four of the arms, escape platforms were placed, and the mice had to learn to associate a specific visual configuration (e.g., black dots on a white background, or horizontal black and white stripes) with the presence or absence of a platform (29). In contrast to the results with the spatial radial arm maze, no significant differences between the two genotypes were observed in the nonspatial radial arm maze, either in regard to the overall number of errors (Fig. 7B) or in regard to the error rate as a function of memory load (data not shown). Here, too, both +/+ and congenic *Ntan1*$^{-/-}$ mice did not exhibit significant learning over the 11 days of testing. Thus, no conclusions can be drawn about spatial or nonspatial learning. At the same time, it was possible to conclude that

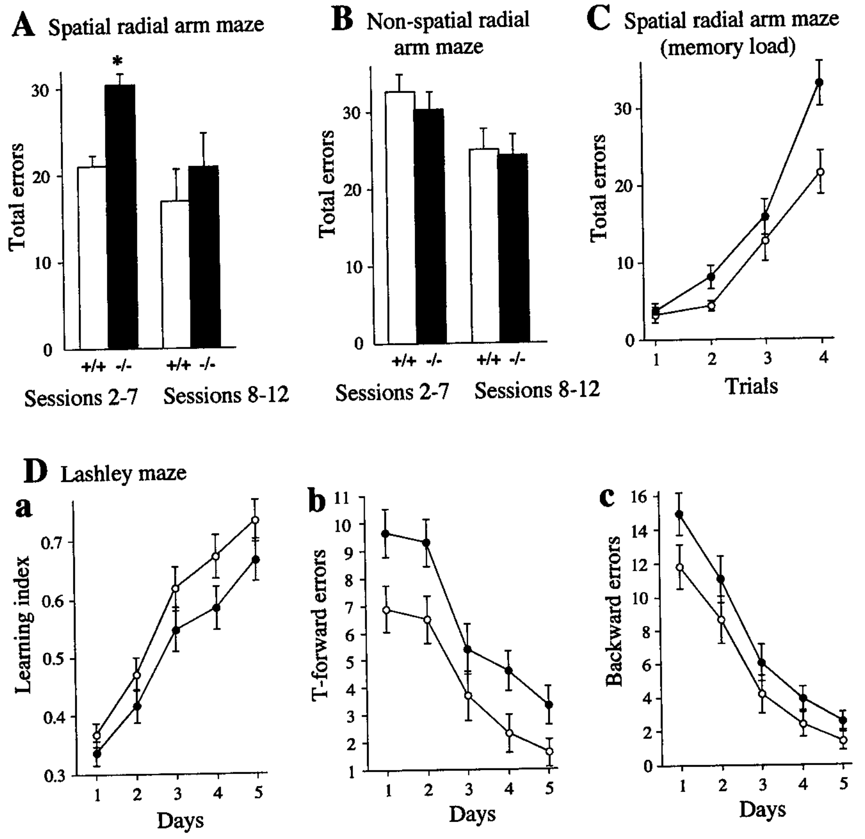


FIG. 7. Spatial learning and memory in *Ntan1*$^{-/-}$ mice. Open symbols and bars represent +/+ mice; solid symbols and bars represent congenic *Ntan1*$^{-/-}$ mice. See the text for details. (A to C) Results for 10 +/+ mice and 10 *Ntan1*$^{-/-}$ mice. (A) Mean daily errors made by mice for each phase in the spatial radial arm maze. The data were separated into those for acquisition phase (sessions 2 to 7) and asymptotic phase (sessions 8 to 12). (B) Same as panel A except that the mice were tested in the nonspatial radial arm maze, in which the spatial cues were replaced by nonspatial cues. (C) Effects of memory load in the water version of the spatial radial arm maze. Shown are total errors per trial. (D) Performance of 29 +/+ mice and 30 congenic *Ntan1*$^{-/-}$ mice in the Lashley maze. Shown are the ratio of the number of correct path segments taken to the total number of segments (learning index) (a), errors in T-choices (b), and errors made by moving away from the goal (c).

Ntan1$^{-/-}$ mice were less effective in dealing with spatial information than the congenic +/+ mice.

Another test of spatial learning and memory was the Lashley III maze. This maze contains (i) cul-de-sacs that an animal has to learn to avoid and (ii) T-choices. Both +/+ and *Ntan1*$^{-/-}$ mice were able to learn the maze, but *Ntan1*$^{-/-}$ mice were significantly less competent, as measured by the learning index, T-forward errors (making a wrong T-choice), and the total number of backward errors (Fig. 7D). When retested 7 or 8 weeks later, *Ntan1*$^{-/-}$ mice had lower scores on the first trial on all three measures, showing inferior long-term retention (data not shown). These results were similar to the findings with the Morris maze (see above), except that there was a significant difference in the original learning of the Lashley maze favoring the +/+ mice (Fig. 7D). This difference may stem from the fact that the Lashley maze can be learned using intramaze cues to memorize the pathway of left and right turns, in addition to the use of extramaze cues, which are present in both the Lashley and Morris maze tests. The results of the two radial arm studies (Fig. 7A to C) indicated that the *Ntan1*$^{-/-}$ mice used spatial information less effectively than the +/+ mice. In summary, the results of spatial memory tests (Fig. 7) strongly suggested that mouse Nt^{N}-amidase contributes to the processing of spatial information and to long-term retention of spatial learning.

DISCUSSION

Homologs of the 310-residue mammalian Nt^{N}-amidase, encoded by the *Ntan1* gene (24, 59), are present in other vertebrates, in arthropods such as *D. melanogaster*, and in plants such as *A. thaliana* but are apparently absent from the nematode *Caenorhabditis elegans* (data not shown). The deamidation of N-terminal (and apparently only N-terminal) Asn residue in proteins or short peptides is the only known enzymatic activity of Nt^{N}-amidase. This enzyme was operationally defined as a component of the N-end rule pathway, because it converts, in vivo, the N-terminal Asn of engineered N-end rule substrates into N-terminal Asp, which is then conjugated by R-transferase to Arg, a primary destabilizing residue (34, 63). Physiological substrates of Nt^{N}-amidase remain to be identified. Physiological substrates are also unknown for the fungal Nt-amidase, which can deamidate either N-terminal Asn or N-terminal Gln and lacks significant sequence similarities to metazoan Nt^{N}-amidase (6). In the present work, mouse Nt^{N}-amidase was studied using several approaches, including targeted mutagenesis. We report the following results.

(i) Both copies of the mouse *Ntan1* gene, which encodes Nt^{N}-amidase, were replaced by a deletion/disruption allele. The resulting *Ntan1*$^{-/-}$ mice were viable and fertile. Their behavioral phenotypes are described below.

(ii) EF cells from *Ntan1*$^{-/-}$ mice lacked Nt^{N}-amidase activity, in contrast to congenic +/+ EF cells. In addition, among the normally short-lived substrates of the N-end rule pathway, only those bearing N-terminal Asn, a tertiary destabilizing residue (see the introduction), became long-lived in the *Ntan1*$^{-/-}$ EF cells. Thus, at least these cells, and by inference the *Ntan1*$^{-/-}$ mice, lacked the asparagine branch of the N-end rule pathway.

(iii) The *Ntan1*$^{-/-}$ EF cells contained wild-type levels of Nt^{Q}-amidase, which mediates the activity of N-terminal Gln, another tertiary destabilizing residue in the N-end rule. Thus, Nt^{N}-amidase and Nt^{Q}-amidase (the latter enzyme remains to be isolated and characterized) are encoded by different genes.

(iv) The brains and testes of +/+ mice contained, respectively, the 1.4- and 1.6-kb *Ntan1* mRNAs. In addition, the testes but not the brains of +/+ mice also contained the *Ntan1*-derived RNAs of 1.1 and 4 kb. The brains and testes of *Ntan1*$^{-/-}$ mice lacked, respectively, the 1.4- and 1.6-kb *Ntan1* mRNAs. However, the testes of *Ntan1*$^{-/-}$ mice retained the testis-specific 1.1- and 4-kb *Ntan1*-derived RNAs, which were found to lack the 5′-half of the *Ntan1* ORF (including the region deleted in *Ntan1*$^{-/-}$ mice) and to encompass the 3′ half of *Ntan1* ORF. Whether these testis-specific transcripts are physiologically relevant in +/+ mice remains to be determined.

(v) Deletion-disruption of *Ntan1* in the *Ntan1*$^{-/-}$ mice did not significantly change the levels of mRNAs encoding other components of the N-end rule pathway, such as ATE1p (R-transferase), UBR1p (E3α, the pathway's E3 component), UBR2p (a homolog of UBR1p), and mHR6Bp (E2$_{14K}$).

(vi) *Ntan1*, which is expressed, at different levels, in most if not all tissues of adult mice (24), was found through in situ hybridization to be particularly strongly expressed in the branchial arches, the tail bud, and the limb buds of E10 embryos. This expression pattern was indistinguishable from that of *Ubr1* (36), consistent with the assignment of NTAN1p (Nt^{N}-amidase) and UBR1p (E3α) to the same pathway.

(vii) An NTAN1p-GFP fusion was enriched in the nucleus of transfected mouse 3T3 cells, in contrast to free GFP, which was uniformly distributed. NTAN1p-GFP was also present in the nucleus-proximal cytoplasm but was apparently much less abundant at the cells' periphery. This localization pattern of the mammalian Nt^{N}-amidase is quite different from that of the *S. cerevisiae NTA1*-encoded Nt-amidase, most of which is located in the yeast mitochondria (H. R. Wang and A. Varshavsky, unpublished data). In contrast to the mammalian Nt^{N}-amidase, the yeast Nt-amidase, which has no significant sequence similarities to the metazoan enzyme, can deamidate either N-terminal Asn or N-terminal Gln (6).

(viii) While the abnormal phenotypes of *Ntan1*$^{-/-}$ mice are apparently not confined to behavioral/cognitive alterations (see Results), it is the latter that have been analyzed in some detail. Briefly, the *Ntan1*$^{-/-}$ mice were indistinguishable from congenic +/+ mice in motor coordination and general physical performance but had reduced spontaneous activity and less effective spatial memory. Remarkably, the exploratory behavior of *Ntan1*$^{-/-}$ mice was found to be particularly different from that of congenic +/+ mice in the presence of social interactions. If a previously untested *Ntan1*$^{-/-}$ mouse was placed on a small, slightly elevated platform with its +/+ littermate (which it grew up with in the same cage), the *Ntan1*$^{-/-}$ mouse left the platform much more slowly than the +/+ mouse. In contrast, when an *Ntan1*$^{-/-}$ mouse was subjected to this test in the presence of a congenic nonlittermate +/+ mouse, it tended to leave the platform faster than the +/+ mouse, even though *Ntan1*$^{-/-}$ mice exhibited diminished exploratory activity when tested alone (Fig. 6).

Our working hypothesis is that *Ntan1*$^{-/-}$ mice are socially recessive, in relation to congenic +/+ mice, and in addition have a lower spontaneous exploratory activity. In a relatively unstressful setting, such as when it is placed on the platform side-by-side with a (familiar) littermate, the behavior of *Ntan1*$^{-/-}$ mice is governed largely by their inherently lower exploratory activity, yielding the observation that these mice leave the platform much more slowly than their +/+ littermates (Fig. 6A). By contrast, in an otherwise identical test with a +/+ nonlittermate, the postulated social recessiveness of *Ntan1*$^{-/-}$ mice becomes their behavior-governing trait. As a result, they leave the platform faster than the +/+ nonlittermates, to reduce the proximity-induced anxiety (Fig. 6B).

The socially conditioned exploratory phenotype of *Ntan1*$^{-/-}$ mice, identified through the platform-leaving test (Fig. 6), has not been previously described, to our knowledge, with other mouse strains. It is difficult to compare this finding with behavioral studies of other mouse mutants, because most of the earlier analyses of exploratory behavior used solitary mice. An example of differential effect of stress on a specific mutant is provided by the analysis of mice lacking the serotonin receptor 1A. These mice exhibited decreased exploratory activity and increased fear of aversive environments (in comparison to +/+ mice), as measured by the open-field test and the elevated plus-maze test, respectively. However, the mutant mice were indistinguishable from +/+ littermates in a less stressful setting such as their home cages (50).

A parsimonious molecular interpretation of our results (Fig. 5 to 7) is that a normally short-lived regulatory protein(s) that is targeted for degradation by the N-end rule pathway through the protein's N-terminal Asn becomes long-lived in *Ntan1*$^{-/-}$ mice. The resulting increase in the steady-state concentration of this protein(s) alters the functioning of relevant neural networks in the adult brain or, nonalternatively, changes these networks in the course of their establishment during development. That the inactivation of a Ub-dependent proteolytic pathway could affect cognitive functions is illustrated, for example, by the identification of *UBE3A* as a human gene encoding an E3 protein of the Ub system called E6AP. This gene is mutated in the Angelman's syndrome, the symptoms of which include motor dysfunction and mental retardation (33, 45). Mice lacking *UBE3A* exhibit deficits in contextual learning and long-term potentiation (31).

The absence of severe impairments in the *Ntan1*$^{-/-}$ mice should make them particularly suitable as vehicles for the approaches to conditional mutagenesis that use conditional destabilization of a protein of interest. One version of this approach could use a transgenic mouse strain that lacks endogenous *Ntan1* and expresses this gene from one of the previously constructed promoters whose activity can be controlled by small molecules such as tetracycline or ecdysone (9). If a protein of interest is modified, through knock-in mutagenesis and the Ub fusion technique (63), to bear an Asn-containing N-degron, the resulting protein would be long-lived in the absence of the *Ntan1*-encoded Nt^{N}-amidase but short-lived, and therefore scarce, in the presence of Nt^{N}-amidase. In this method, the steady-state level, and hence the activity, of a protein of interest is controlled through *Ntan1*-dependent, regulated changes of the protein's metabolic stability. One advantage of this strategy is that it does not require alterations of a promoter that expresses a gene of interest, thereby avoiding potential perturbations of the promoter regulation during development and differentiation. This approach may be combined with the existing technologies for conditional mutagenesis of the mouse (13, 55), enhancing them through regulated degradation of a protein of interest.

Inasmuch as behavior is an emergent property of highly complex neural networks interacting with the musculoskeletal apparatus, a behavioral alteration that is caused by the absence of a specific protein from the brain is difficult to understand in terms of higher-order neural events even in the case of a protein (e.g., a neurotransmitter-gated ion channel) whose molecular function in the individual neurons is clear. The difficulty is further increased in the case of a protein such as Nt^N-amidase, whose physiological substrates remain to be identified. At the same time, the absence of a priori reasons to suspect a specific role for Nt^N-amidase in the brain's functions and the robustness of behavioral differences between the $Ntan1^{-/-}$ and congenic +/+ mice (Fig. 5 to 7) make these findings particularly intriguing. Further advances in this inquiry will require, at minimum, the identification of physiological substrates of Nt^N-amidase. The availability of $Ntan1^{-/-}$ mice and cell lines derived from them should facilitate the discovery of these substrates.

ACKNOWLEDGMENTS

We are grateful to members of the Caltech Transgenic Facility, especially S. Pease, B. Kennedy, and A. Granados, for the care of mice and expert technical help. We thank S. Grigoryev for helpful discussions at the beginning of this study, N. Barteneva for assistance with some of the early experiments, S. Offermanns for advice and help with preparation of embryonic fibroblasts, and R. C. Mulligan and R. Jaenisch for gifts of plasmids.

This work was supported by grants GM31530 and DK39520 from the National Institutes of Health to A.V.

REFERENCES

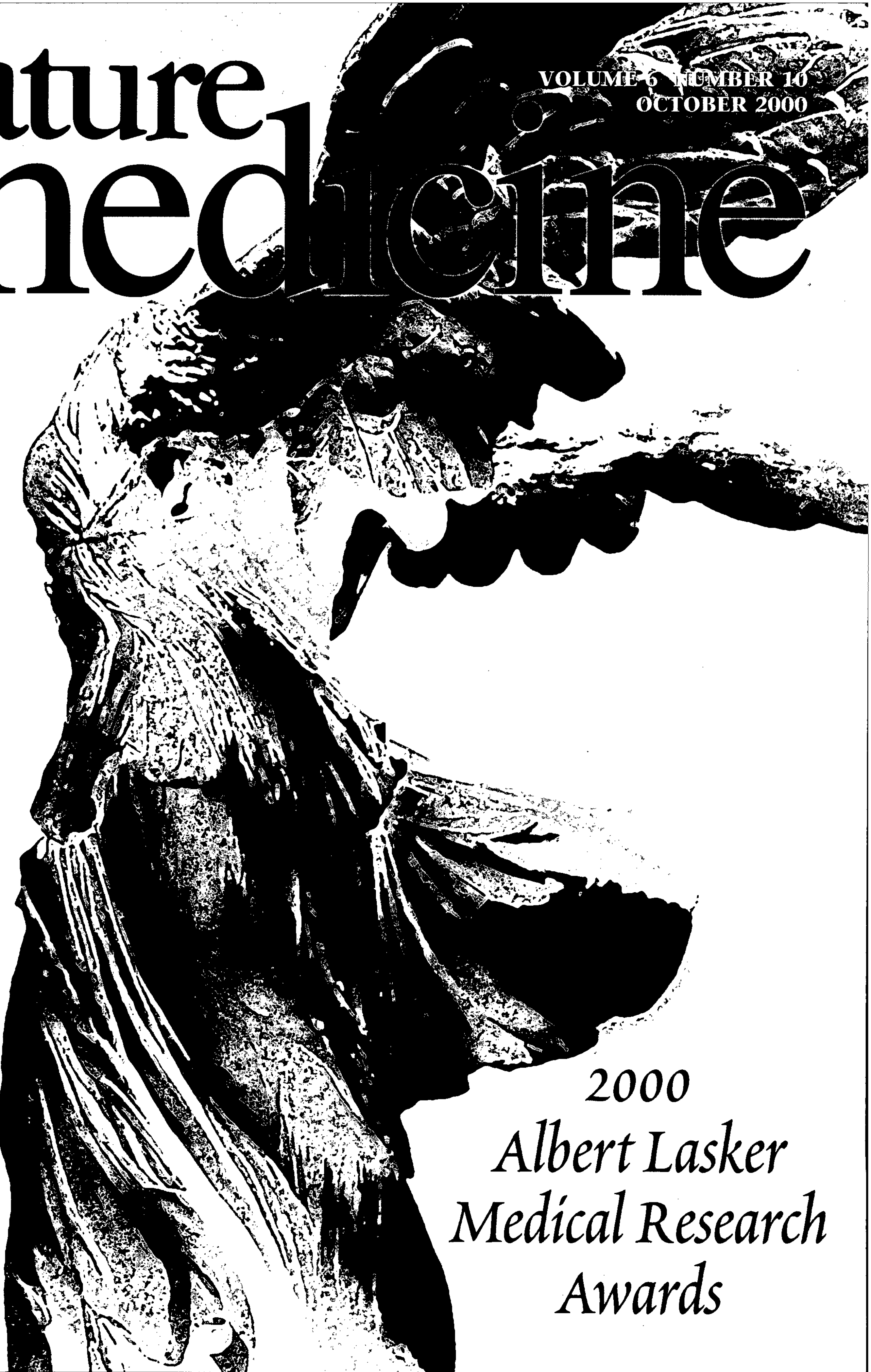
1. **Alagramam, K., F. Naider, and J. M. Becker.** 1995. A recognition component of the ubiquitin system is required for peptide transport in *Saccharomyces cerevisiae*. Mol. Microbiol. **15:**225–234.
2. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. Struhl (ed.).** 1998. Current protocols in molecular biology. Wiley-Interscience, New York, N.Y.
3. **Bachmair, A., D. Finley, and A. Varshavsky.** 1986. *In vivo* half-life of a protein is a function of its amino-terminal residue. Science **234:**179–186.
4. **Bachmair, A., and A. Varshavsky.** 1989. The degradation signal in a short-lived protein. Cell **56:**1019–1032.
5. **Baker, R. T., and A. Varshavsky.** 1991. Inhibition of the N-end rule pathway in living cells. Proc. Natl. Acad. Sci. USA **87:**2374–2378.
6. **Baker, R. T., and A. Varshavsky.** 1995. Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway. J. Biol. Chem. **270:** 12065–12074.
7. **Balogh, S. A., C. S. McDowell, A. Stavnezer, and V. H. Denenberg.** 1999. A behavioral and neuroanatomical assessment of an inbred substrain of 129 mice, with behavioral comparisons to C57BL/6J mice. Brain Res. **836:**38–48.
8. **Balzi, E., M. Choder, W. Chen, A. Varshavsky, and A. Goffeau.** 1990. Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. J. Biol. Chem. **265:**7464–7471.
9. **Baron, U., D. Schnappinger, V. Helbl, M. Gossen, W. Hillen, and H. Bujard.** 1999. Generation of conditional mutants in higher eukaryotes by switching between the expression of two genes. Proc. Natl. Acad. Sci. USA **96:**1013–1018.
10. **Baumeister, W., J. Walz, F. Zühl, and E. Seemüller.** 1998. The proteasome: paradigm of a self-compartmentalizing protease. Cell **92:**367–380.
11. **Beaudin, S., and R. Lalonde.** 1997. The effects of pentobarbital on spatial learning, motor coordination, and exploration. Pharmacol. Biochem. Behav. **57:**111–114.
12. **Byrd, C., G. C. Turner, and A. Varshavsky.** 1998. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. EMBO J. **17:**269–277.
13. **Capecchi, M. R.** 1989. Altering the genome by homologous recombination. Science **244:**1288–1292.
14. **Chang, S. E., J. Keen, E. B. Lane, and J. Taylor-Papadimitriou.** 1982. Establishment and characterization of SV40-transformed human breast epithelial cell lines. Cancer Res. **42:**2040–2053.
15. **Chau, V., J. W. Tobias, A. Bachmair, D. Marriott, D. J. Ecker, D. K. Gonda, and A. Varshavsky.** 1989. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science **243:**1576–1583.
16. **Conlon, R. A., and J. Rossant.** 1992. Exogenous retinoic acid rapidly induces anterior ectopic expression of murine Hox-2 genes *in vivo*. Development **116:**357–368.
17. **Coux, O., K. Tanaka, and A. L. Goldberg.** 1996. Structure and functions of the 20S and 26S proteasomes. Annu. Rev. Biochem. **65:**801–817.
18. **Davydov, I. V., D. Patra, and A. Varshavsky.** 1998. The N-end rule pathway in *Xenopus* egg extracts. Arch. Biochem. Biophys. **357:**317–325.
19. **deGroot, R. J., T. Rümenapf, R. J. Kuhn, and J. H. Strauss.** 1991. Sindbis virus RNA polymerase is degraded by the N-end rule pathway. Proc. Natl. Acad. Sci. USA **88:**8967–8971.
20. **DeMartino, G. N., and C. A. Slaughter.** 1999. The proteasome, a novel protease regulated by multiple mechanisms. J. Biol. Chem. **274:**22123–22126.
21. **Denenberg, V. H., N. Talgo, D. A. Carroll, S. Freter, and R. Deni.** 1991. A computer-aided procedure for measuring Lashley III maze performance. Physiol. Behav. **50:**857–861.
22. **Denenberg, V. H., N. W. Talgo, N. S. Waters, and G. H. Kenner.** 1990. A computer-aided procedure for measuring swim rotation. Physiol. Behav. **47:**1023–1025.
23. **Gonda, D. K., A. Bachmair, I. Wünning, J. W. Tobias, W. S. Lane, and A. Varshavsky.** 1989. Universality and structure of the N-end rule. J. Biol. Chem. **264:**16700–16712.
24. **Grigoryev, S., A. E. Stewart, Y. T. Kwon, S. M. Arfin, R. A. Bradshaw, N. A. Jenkins, N. G. Copeland, and A. Varshavsky.** 1996. A mouse amidase specific for N-terminal asparagine. The gene, the enzyme, and their function in the N-end rule pathway. J. Biol. Chem. **271:**28521–28532.
25. **Haas, A. J., and T. J. Siepman.** 1997. Pathways of ubiquitin conjugation. FASEB J. **11:**1257–1268.
26. **Hershko, A., and A. Ciechanover.** 1998. The ubiquitin system. Annu. Rev. Biochem. **76:**425–479.
27. **Hochstrasser, M.** 1996. Ubiquitin-dependent protein degradation. Annu. Rev. Genet. **30:**405–439.
28. **Hondermarck, H., J. Sy, R. A. Bradshaw, and S. M. Arfin.** 1992. Dipeptide inhibitors of ubiquitin-mediated protein turnover prevent growth factor-induced neurite outgrowth in rat pheochromocytoma PC12 cells. Biochem. Biophys. Res. Commun. **30:**280–288.
29. **Hyde, L. A., and V. H. Denenberg.** 1999. BXSB mice can learn complex pattern discrimination. Physiol. Behav. **66:**437–439.
30. **Hyde, L. A., B. J. Hoplight, and V. H. Denenberg.** 1998. Water version of the radial-arm maze: learning in three inbred strains of mice. Brain Res. **785:** 236–244.
31. **Jiang, Y. H., D. Armstrong, U. Albrecht, C. M. Atkins, J. L. Noebels, G. Eichele, J. D. Sweatt, and A. L. Beaudet.** 1998. Mutation of the Angelman ubiquitin ligase in mice causes increased cytoplasmic p53 and deficits of contextual learning and long-term potentiation. Neuron **21:**799–811.
32. **Johnson, E. S., D. K. Gonda, and A. Varshavsky.** 1990. *Cis-trans* recognition and subunit-specific degradation of short-lived proteins. Nature **346:**287–291.
33. **Kishino, T., M. Lalande, and J. Wagstaff.** 1997. UBE3A/E6-AP mutations cause Angelman syndrome. Nat. Genet. **15:**70–73.
34. **Kwon, Y. T., A. S. Kashina, and A. Varshavsky.** 1999. Alternative splicing results in differential expression, activity, and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway. Mol. Cell. Biol. **19:**182–193.
35. **Kwon, Y. T., F. Lévy, and A. Varshavsky.** 1999. Bivalent inhibitor of the N-end rule pathway. J. Biol. Chem. **274:**18135–18139.
36. **Kwon, Y. T., Y. Reiss, V. A. Fried, A. Hershko, J. K. Yoon, D. K. Gonda, P. Sangan, N. G. Copeland, N. A. Jenkins, and A. Varshavsky.** 1998. The mouse and human genes encoding the recognition component of the N-end rule pathway. Proc. Natl. Acad. Sci. USA **95:**7898–7903.
37. **Laney, J. D., and M. Hochstrasser.** 1999. Substrate targeting in the ubiquitin system. Cell **97:**427–430.
38. **Lawson, T. G., D. L. Gronros, P. E. Evans, M. C. Bastien, K. M. Michalewich, J. K. Clark, J. H. Edmonds, K. H. Graber, J. A. Werner, B. A. Lurvey, and J. M. Cate.** 1999. Identification and characterization of a protein destruction signal in the encephalomyocarditis virus 3C protease. J. Biol. Chem. **274:**9871–9880.
39. **Lecker, S. H., V. Solomon, S. R. Price, Y. T. Kwon, W. E. Mitch, and A. L. Goldberg.** 1999. Ubiquitin conjugation by the N-end rule pathway and mRNAs for its components increase in muscles of diabetic rats. J. Clin. Investig. **104:**1411–1420.
40. **Lévy, F., N. Johnsson, T. Rumenapf, and A. Varshavsky.** 1996. Using ubiquitin to follow the metabolic fate of a protein. Proc. Natl. Acad. Sci. USA **93:**4907–4912.
41. **Li, J., and C. Pickart.** 1995. Inactivation of arginyl-tRNA protein transferase by a bifunctional arsenoxide: identification of residues proximal to arsenoxide site. Biochemistry **34:**139–147.
42. **Lister, R. G.** 1987. The use of a plus-maze to measure anxiety in the mouse. Psychopharmacology **92:**180–185.
43. **Madura, K., and A. Varshavsky.** 1994. Degradation of Gα by the N-end rule pathway. Science **265:**1454–1458.
44. **Maniatis, T.** 1999. A ubiquitin ligase complex essential for the NF-κB, Wnt/Wingless, and Hedgehog signaling pathways. Genes Dev. **13:**505–510.
45. **Matsuura, T., J. S. Sutcliffe, P. Fang, R. J. Galjaard, Y. H. Jiang, C. S. Benton, J. M. Rommens, and A. L. Beaudet.** 1997. *De novo* truncating

mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. Nat. Genet. **15**:74–77.

46. **Morris, R.** 1984. Development of a water-maze procedure for studying spatial learning in the rat. J. Neurosci. Methods **11**:47–60.
47. **Ota, I. M., and A. Varshavsky.** 1993. A yeast protein similar to bacterial two-component regulators. Science **262**:566–569.
48. **Peters, J.-M., R. W. King, and R. J. Deshaies.** 1998. Cell cycle control by ubiquitin-dependent proteolysis, p. 345–387. *In* J. M. Peters, J. R. Harris, and D. Finley (ed.), Ubiquitin and the biology of the cell. Plenum Press, New York, N.Y.
49. **Pickart, C. M.** 1997. Targeting of substrates to the 26S proteasome. FASEB J. **11**:1055–1066.
50. **Ramboz, S., R. Oosting, D. A. Amara, H. F. Kung, P. Blier, M. Mendelsohn, J. J. Mann, D. Brunner, and R. Hen.** 1998. Serotonin receptor 1A knockout: an animal model of anxiety-related disorder. Proc. Natl. Acad. Sci. USA **95**:14476–14481.
51. **Rechsteiner, M.** 1998. The 26S proteasome, p. 147–189. *In* J. M. Peters, J. R. Harris, and D. Finley (ed.), Ubiquitin and the biology of the cell. Plenum Press, New York, N.Y.
52. **Robertson, E. J.** 1987. Embryo-derived stem cell lines, p. 71–112. *In* E. J. Robertson (ed.), Teratocarcinomas and embryonic stem cells: a practical approach. IRL Press, Oxford, United Kingdom.
53. **Royce, J. R.** 1972. Avoidance conditioning in nine strains of inbred mice using optimal stimulus parameters. Behav. Genet. **2**:107–110.
54. **Schauber, C., L. Chen, P. Tongaonkar, I. Vega, and K. Madura.** 1998. Sequence elements that contribute to the degradation of yeast G-alpha. Genes Cells **3**:307–319.
55. **Schwenk, F., R. Kuhn, P. O. Angrand, K. Rajewsky, and A. F. Stewart.** 1998. Temporally and spatially regulated somatic mutagenesis in mice. Nucleic Acids Res. **26**:1427–1432.
56. **Sijts, A. J., I. Pilip, and E. G. Pamer.** 1997. The Listeria monocytogenes-secreted p60 protein is an N-end rule substrate in the cytosol of infected cells. Implications for major histocompatibility complex class I antigen processing of bacterial proteins. J. Biol. Chem. **272**:19261–19268.
57. **Solomon, V., V. Baracos, P. Sarraf, and A. Goldberg.** 1998. Rates of ubiquitin conjugation increase when muscles atrophy, largely through activation of the N-end rule pathway. Proc. Natl. Acad. Sci. USA **95**:12602–12607.
58. **Stewart, A.** 1995. Trends in genetics nomenclature guide. Elsevier Science, Ltd., Cambridge, United Kingdom.
59. **Stewart, A. E., S. M. Arfin, and R. A. Bradshaw.** 1995. The sequence of porcine protein NH2-terminal asparagine amidohydrolase. A new component of the N-end rule pathway. J. Biol. Chem. **270**:25–28.
60. **Suzuki, T., and A. Varshavsky.** 1999. Degradation signals in the lysine-asparagine sequence space. EMBO J. **18**:6017–6026.
61. **Taban, C. H., H. Hondermarck, R. A. Bradshaw, and B. Boilly.** 1996. Effect of a dipeptide inhibiting ubiquitin-mediated protein degradation on nerve-dependent limb regeneration in the newt. Experientia **52**:865–870.
62. **Tobias, J. W., and A. Varshavsky.** 1991. Cloning and functional analysis of the ubiquitin-specific protease gene *UBP1* of *Saccharomyces cerevisiae*. J. Biol. Chem. **266**:12021–12028.
63. **Varshavsky, A.** 1996. The N-end rule: functions, mysteries, uses. Proc. Natl. Acad. Sci. USA **93**:12142–12149.
64. **Varshavsky, A.** 1997. The ubiquitin system. Trends Biochem. Sci. **22**:383–387.
65. **Wang, Y. M., and N. A. Ingoglia.** 1997. N-terminal arginylation of sciatic nerve and brain proteins following injury. Neurochem. Res. **22**:1453–1459.
66. **Wilkinson, K., and M. Hochstrasser.** 1998. The deubiquitinating enzymes, p. 99–126. *In* J.-M. Peters, J. R. Harris, and D. Finley (ed.), Ubiquitin and the biology of the cell. Plenum Press, New York, N.Y.

nature medicine

VOLUME 6 NUMBER 10
OCTOBER 2000



2000 Albert Lasker Medical Research Awards

nature medicine

EDITORIAL OFFICE **medicine@natureny.com**

345 Park Avenue South, New York, NY 10010-1707
Tel: (212) 726 9325, Fax: (212) 683 5751
Editor: Beatrice Renault
Senior Editor: Robert Frederickson
Senior Editor, News: Karen Birmingham
Associate Editor, News and Views: Kristine Novak
Assistant Editor, Research manuscripts: Leila Alland
Assistant Editor, Research manuscripts: Simon Noble
Assistant Editor, New Technology: Diane Gershon
Production Editor: Stephen Horwitz
Editorial Assistant: Kelly Ann Hamilton
Copy Editor: Jennifer Fosmire

MANAGEMENT OFFICES

Nature America Inc.
345 Park Avenue South, New York, NY 10010-1707
Tel: (212) 726 9200, Fax: (212) 696 9006
Publisher: Adrian J. Ivinson
Vice President of Advertising: Marion Delaney
Vice President of Marketing: Phillip LoFaso
Senior Marketing Manager: Sara Merrill Girard
Circulation Manager: Richard Chung
Creative Director: Lewis E. Long
Art Director: Christie Albin
Production Systems Manager: Christine C. Bretherton
Production Coordinator: Diane Rios
Ad Traffic Coordinator: Renee Burczyk

Nature Publishing Group
Porters South, Crinan Street, London N1 9XW
Tel: 44 207 833 4000, Fax: 44 207 843 4996
Managing Director: Stefan von Holtzbrinck
Editor-in-Chief, Nature *Publications*: Philip Campbell
International Projects Manager: Liz Allen
Group Marketing Manager: Michelle Tempest-Mitchell
Marketing Executive: Timothy Redding

Nature Japan K.K.
Shin-Mitsuke Building 3F, 3-6 Ichigaya Tamachi, Shinjuku-ku, Tokyo 162-0843
Tel: 81 3 3267 8751, Fax: 81 3 3267 8746
Asia-Pacific Publishing Director: David Swinbanks
Manager: Koichi Nakamura
Marketing Manager: Michael Thompson
Marketing Executive: Danny Choo

DISPLAY ADVERTISING **display@natureny.com (US/Canada)**
display@nature.com (Europe)

US Sales Director: William G. Moran, Tel: (212) 726 9243, Fax: (212) 696 9481
European Sales Manager: Rachel Burley, Tel: 44 207 843 4960, Fax: 44 207 843 4996
Japan Sales Manager: Kate Cowan, Tel: 81 3 3267 8751, Fax: 81 3 3267 8746

Display sales offices
US, East & Canada: Debbie Mecca, Tel: (212) 726 9334
US, Midwest: Jim Secretario, Tel: (212) 726 9280
US, West & Canada: Lynne Stickrod and Jim Secretario, Tel: (800) 229 9758
Australia: Anne Murray, World Media Network, Tel: 61 2 299 5677, Fax: 61 2 299 6178
Israel: Marcus Sheff, The Word Shop Ltd, Tel: 972 3 575 4426, Fax: 972 3 751 0687
Sweden: Mathias Saving, Andrew Karnig & Assocs, Tel: 46 8 442 7050, Fax: 46 8 442 7059
Switzerland: Verena Lowenthal, Interpress, Tel: 41 52 761 30 30, Fax: 41 52 761 32 00

CLASSIFIED ADVERTISING **classified@natureny.com (US/Canada)**
classified@nature.com (Europe)

US Sales Manager: Ben Crowe, Tel: (212) 726 9245, Fax: (212) 696 9482
International Sales Manager: Fabien Savenay, Tel: 44 207 843 4961, Fax: 44 207 843 4996
European Sales Manager: Leonie Weiss, Tel: 44 207 843 4961, Fax: 44 207 843 4996

CUSTOMER SERVICE **subscriptions@natureny.com (The Americas)**
subscriptions@nature.com (Europe)
subscriptions@naturejpn.com (Japan)
harpal@nature.com (India)

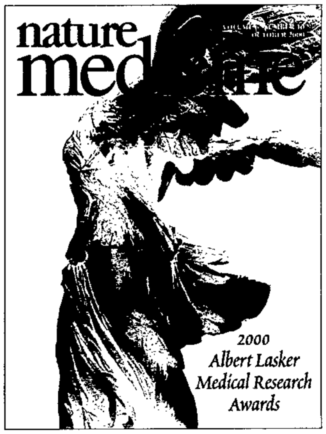
For new subscriptions, renewals, change of address, and back issues:
In the Americas: *Nature Medicine,* Subscription Department, P.O. Box 5054, Brentwood, TN 37024-5054. Tel: (800) 524 0384, Direct Dial (615) 377 3322, Fax: (615) 377 0525.
Outside the Americas: *Nature Medicine*, Macmillan Magazines Ltd., Porters South, Crinan Street, London N1 9XW, UK. Tel: 44 1256 329 242, Fax: 44 1256 812 358.
In Japan: *Nature Medicine*, Nature Japan K.K., Shin-Mitsuke Building 3F, 3-6 Ichigaya Tamachi, Shinjuku-ku, Tokyo 162-0843. Tel: 81 3 3267 8751, Fax: 81 3 3267 8746.
In India: Harpal Singh Gill, Macmillan Magazines Ltd, 5A/12 Ansari Road, Daryaganj, New Delhi, 110 002 India. Tel/Fax: 00 91 11 327 2010

REPRINTS **reprints@natureny.com**

Layonne Holmes, *Nature Medicine* Reprint Department, 345 Park Avenue South, New York, NY 10010-1707, US. Tel: (212) 726 9278, Fax: (212) 679 0843.
To order 600 or more copies in the Americas: Melissa Rocco, Tel: (212) 726 9347, Fax: (212) 696 90769, e-mail: m.rocco@natureny.com.
To order 1,000 or more copies outside the Americas: *Nature Medicine* Reprint Department, Macmillan Magazines Ltd., Porters South, 4-6 Crinan Street, London N1 9XW, UK. Tel: 44 207 843 4987, Fax: 44 207 843 4996, e-mail: reprints@nature.com.

WORLD WIDE WEB **http://medicine.nature.com/**

Director of Electronic Publishing Services: Don Fick
Manager of Structured Data Development: Joe Landolfi
Electronic Production Coordinator: Kelly White



The cover shows the statue known as *The Winged Victory of Samothrace*, the Lasker Foundation's symbol of victory over disability, disease and death.

nature medicine

REPRINTED FROM VOLUME 6, NUMBER 10, OCTOBER 2000 **http://medicine.nature.com**

FOREWORD

From the Editor i
BEATRICE RENAULT

Laskers for 2000: Ubiquitin Pathway, Hepatitis C and Sydney Brenner i
JOSEPH L. GOLDSTEIN

BASIC MEDICAL RESEARCH AWARD

The ubiquitin system iii
AVRAM HERSHKO, AARON CIECHANOVER & ALEXANDER VARSHAVSKY

CLINICAL MEDICAL RESEARCH AWARD

Hepatitis C Virus and eliminating post-transfusion hepatitis xii
HARVEY J. ALTER & MICHAEL HOUGHTON

SPECIAL ACHIEVEMENT IN MEDICAL SCIENCE AWARD

From cell physiology to cell physiology xvii
SYDNEY BRENNER

Nature Medicine (ISSN 1078-8956) is published monthly by Nature America Inc., headquartered at 345 Park Avenue South, New York, NY 10010-1707. **Editorial Office:** 345 Park Avenue South, New York, NY 10010-1707. Tel: (212) 726 9325, Fax: (212) 683 5751. **North American Advertising:** 345 Park Avenue South, New York, NY 10010-1707. Tel: (212) 726 9200, Fax: (212) 696 9006. **European Advertising:** Nature Medicine, Porters South, Crinan Street, London N1 9XW, UK. Tel: 44 207 833 4000, Fax: 44 207 843 4996, Telex 262024. **New subscriptions, renewals, changes of address, back issues and all customer service questions in the Americas** should be addressed to: *Nature Medicine* Subscription Department, P.O. Box 5054, Brentwood, TN 37024-5054. Tel:(800) 524 0384, Direct Dial (615) 377 3322, Fax: (615) 377 0525. Outside the Americas: *Nature Medicine,* Macmillan Magazines Ltd., Houndsmill, Brunel Road, Basingstoke, RG21 6XS, U.K.. Tel: +44-(0)1256-329242. Fax: +44-(0)1256 812358. Email: subscriptions@nature.com. Annual subscription rates: North America US$650 (institutional/corporate), $149 (individual making personal payment; Canada add 7% for GST, BN: 14091 1595 RT); U.K./Europe:£475 (institutional/corporate), £110 (individual making personal payment), £99 (student); Rest of world (excluding Japan): £520 (institutional/corporate), £175 (individual making personal payment) £125(student). Japan: Contact Nature Japan K.K., Shin-Mitsuke Building 3F, 3-6 Ichigaya Tamachi, Shinjuku-ku, Tokyo 162-0843. **Authorization to photocopy** for internal or personal use, or internal or personal use of specific clients, is granted by *Nature Medicine* to libraries and others registered with the Copyright Clearance Center (CCC) Transactional Routing Service, provided the base fee of $9.00 an article (or $1.00 a page) is paid direct to CCC, 27 Congress Street, Salem, MA 01970, US. Periodicals postage paid at New York, NY, 10010, and at additional mailing offices. **POSTMASTER:** Send address changes to *Nature Medicine* Subscription Department, P.O. Box 5054, Brentwood, TN 37024-5054. Executive Officers of Nature America Inc.: Nicholas Byam Shaw, Chairman of the Board; Jan Velterop, President; Edward Valis, Secretary Treasurer. Published in Japan by Nature Japan K.K., Shin-Mitsuke Building 3F, 3-6 Ichigaya Tamachi, Shinjuku-ku, Tokyo 162-0843, Japan. Printed by Cadmus Journal Services, Richmond, VA, US. Copyright ©2000 Nature America Inc.

nature medicine
VOLUME 6 • NUMBER 10 • OCTOBER 2000

When The Lasker Foundation first approached *Nature Medicine* with the idea of working together to present a series of articles from the winners of their prestigious awards, we were immediately enthusiastic. The prospect of working with such celebrated figures from the biomedical research community is one that would fill any scientific editor with glee. This year, as before, we have not been disappointed. The award winners have again shown themselves not only to be exceptional figures in the community but also accomplished authors with great stories to tell. We thank them for their generous contributions to these pages.

Once again the awards acknowledge very special scientists whose contributions have paved new paths of research, significantly advancing our understanding of diseases and our quest for prevention and treatment. (See Joseph Goldstein's Foreword for a synopsis of their remarkable contributions to the community.) We hope their articles will inspire scientists and non-scientist alike to redouble their commitment to biomedical research and to the ultimate goal of eradicating disease.

BEATRICE RENAULT

Laskers for 2000: Ubiquitin Pathway, Hepatitis C, and Sydney Brenner

The Lasker Award in Basic Medical Research is given each year to honor a fundamental discovery that opens up a new area of biomedical science. This year's recipients—Aaron Ciechanover, Avram Hershko and Alexander Varshavsky—discovered and recognized the broad significance of the ubiquitin system of regulated protein degradation. Ciechanover and Hershko are biochemists at the Technion–Israel Institute of Technology in Haifa. Varshavsky is a molecular biologist at the California Institute of Technology in Pasadena.

The discovery of the ubiquitin system is a 'triple-A' story befitting the three protagonists. When they began their work in 1977–1978, the genetic code had been deciphered, the mechanism of protein synthesis had been delineated and the technologies for cloning and sequencing genes had been discovered. But almost nothing was known about the mechanism of intracellular protein degradation. Then in 1978, Hershko and Ciechanover, working with reticulocyte extracts, discovered that proteins could be targeted for degradation by the covalent attachment of a 76-amino-acid protein, later identified as ubiquitin. In a series of elegant biochemical papers over the next 5 years, they showed that this unprecedented type of protein attachment required the action of three enzymes that sequentially activate (E1), conjugate (E2) and ligate (E3) ubiquitin to protein substrates. They also discovered that ubiquitin–protein conjugates are degraded by an ATP-dependent protease, now known to be the multi-subunit 26S proteasome.

The first evidence that this pathway had a regulatory function in living cells was provided in 1984 by Varshavsky, who traced the cell cycle arrest phenotype in a mouse cell line to a cessation of the degradation of rapidly turning-over proteins. In collaboration with Ciechanover, Varshavsky and Daniel Finley showed that this phenotype was caused by a temperature-sensitive mutation in the ubiquitin-activating enzyme (E1), genetically 'cementing' the biochemical model of Hershko and Ciechanover. Varshavsky went on to establish a genetic system in yeast that showed the first set of rules that specify which proteins undergo ubiquitination.

The physiological importance of the ubiquitin system has now been established for many cellular events involving the cell cycle, antigen processing, inflammation, gene transcription and oncogenesis. Its discovery is reminiscent of the discovery of protein phosphorylation. When Fischer and Krebs first reported protein phosphorylation in the mid-1950s, it was thought to be limited to one or two enzymes in glycogen metabolism. Later work demonstrated its universality, and today we recognize that phosphorylation regulates nearly all metabolic pathways. The same seems to be true for the ubiquitin-mediated proteolytic pathway.

The Lasker Clinical Medical Research Award is given each year to honor a scientific contribution that has improved the treatment of patients in alleviating or eliminating a major medical disease. This year's recipients—Harvey Alter of the

National Institutes of Health and Michael Houghton, leader of a scientific team from Chiron Corp.—did pioneering work that led to the discovery of the hepatitis C virus and the development of screening methods that reduced the risk of blood transfusion-associated hepatitis in the US from 30% in 1970 to nearly zero in 2000.

Harvey Alter has studied the biology of transfusion hepatitis for 40 years, during which time he acquired a priceless collection of 'pedigreed' blood samples—blood collected from the donor of a transfusion and from the donor's recipient. In the mid-1970s he and his collaborators at the National Institutes of Health, Robert Purcell and Stephen Feinstone, used these 'pedigreed' blood samples to show that transfusion hepatitis was not caused by either the hepatitis A virus or the hepatitis B virus, but by some unidentified virus (or viruses), which they called non-A non-B. In studying the natural history of non-A non-B hepatitis, Alter showed that 20% of infected people went on to develop cirrhosis, a serious liver disease. In an attempt to isolate the putative virus, Alter and Purcell injected chimpanzees with samples from patients with non-A non-B hepatitis and demonstrated that the injected animals developed hepatitis.

Despite the early success in transmitting the non-A non-B virus to animals, a conclusive identification proved elusive until 1985. Here the story shifts to Michael Houghton's laboratory at the Chiron Corp., where Houghton and his colleagues Qui-Lim Choo and George Kuo embarked on a risky strategy aimed at molecularly cloning the genome of the putative virus. Starting with infected chimpanzee tissue provided by Daniel Bradley of the Centers for Disease Control, the Chiron team essentially did a 'blind cloning'. They made cDNA libraries from the chimpanzee tissues, introduced them into bacteria and screened for expressed proteins that bound hypothetical antibodies from patients infected with non-A non-B hepatitis but not from uninfected individuals. After 2 years of screening tens of millions of bacterial clones, they eventually identified one clone that encoded a fragment of the genome of an RNA virus belonging to the Flaviviradae family, now known to be the hepatitis C virus. Not only did Houghton, Choo and Kuo succeed in identifying the hepatitis C virus, but they also demonstrated the power of 'blind cloning' as a method to identify an infectious agent that had never been grown in the laboratory, never been visualized microscopically and never been identified serologically. If ever there were a technical *tour de force* of molecular cloning, this was it!

The *pièce de résistance* of the story came in 1989 when the Chiron team and Alter collaborated, using Chiron's new serological reagents and Alter's 'pedigreed' blood samples, to demonstrate that the new hepatitis C virus was indeed the cause of most cases of non-A non-B hepatitis. Since 1990, reliable tests for screening the blood supply for hepatitis C have been in routine use throughout the world.

The Lasker Special Achievement Award in Medical Science is a new award that was inaugurated in 1994 and is given periodically to honor a scientist whose lifetime contributions to biomedical research are universally admired and respected for their creativity, importance and impact. Previous recipients have included Maclyn McCarty (1994), Paul C. Zamecnik (1996), Victor A. McKusick (1997), Daniel E. Koshland, Jr. (1998) and Seymour Kety (1999).

The recipient of this year's award is Sydney Brenner. Brenner is cited for 50 years of brilliant creativity in biomedical science, exemplified by his legendary work on the genetic code; his daring introduction of the roundworm *Caenorhabditis elegans* as a system for tracing the birth and death of every cell in a living animal; his rational voice in the debate on recombinant DNA; and his trenchant wit. Additional information on Brenner is available in *Remarks from the 2000 Lasker Awards Ceremony* (http://www.laskerfoundation.com/library/2000/remarks3.html).

JOSEPH L. GOLDSTEIN
Chair, Lasker Awards Jury

Basic Medical Research Award

The ubiquitin system

AVRAM HERSHKO, AARON CIECHANOVER & ALEXANDER VARSHAVSKY

Ubiquitin-mediated protein degradation: The early days

AVRAM HERSHKO

It has been often stated that until recently the ubiquitin system was thought to be mainly a 'garbage disposal' for the removal of abnormal or damaged proteins. This statement is certainly not true for those who have been interested in the selective and regulated degradation of proteins in cells. The dynamic turnover of cellular proteins was discovered in the pioneering studies of Rudolf Schoenheimer in the 1930s, when he first used isotopically labeled compounds for biological studies[1]. Between 1960 and1970 it became evident that protein degradation in animal cells is highly selective, and is important in the control of specific enzyme concentrations[2]. The molecular mechanisms responsible for this process, however, remained unknown. Some imaginative models have been proposed to account for the selectivity of protein degradation, such as one suggesting that all cellular proteins are rapidly engulfed into the lysosome, but only short-lived proteins are degraded in the lysosome, whereas long-lived proteins escape back to the cytosol[3].

I became interested in the mechanisms of intracellular protein breakdown when I was a post-doctoral fellow in the laboratory of Gordon Tomkins 30 years ago (1969–1971). At that time, the main subject in that laboratory was the mechanism by which corticosteroid hormones cause the increased synthesis of the enzyme tyrosine aminotransferase. I found this subject a bit crowded, so I chose to study a different process that also regulates tyrosine aminotransferase concentration: the degradation of this enzyme. I found that the degradation of tyrosine aminotransferase in cultured hepatoma cells is completely arrested by inhibitors of cellular energy production, such as fluoride or azide[4]. These results confirmed and extended the previous observations of Simpson on the energy dependence of the release of amino acids from liver slices[5]. Similar energy requirements for the degradation of many other cellular proteins were subsequently found in a variety of experimental systems[6].

I was very impressed by the energy dependence of intracellular protein breakdown, because proteolysis itself is an exergonic process that does not require energy. I assumed that there was an as-yet-unknown proteolytic system that uses energy for the highly selective degradation of proteins. After returning to Israel in 1971 and setting up my laboratory at the Technion, my main goal was to identify the energy-dependent system responsible for the degradation of cellular proteins. It took a bit of faith to base my entire research project on the effects of energy 'poisons', because these inhibitors could affect protein breakdown rather indirectly. For example, I remember that when Racker, a great biochemist, visited my laboratory in Haifa in the mid-1970s, he dismissed these observations as being secondary to the inhibition of the proton pump, which maintains the acidic environment in lysosomes. I was convinced, however, that lysosomal autophagy cannot account for the selectivity and regulation of intracellular protein breakdown. I was also convinced that the best way to identify a new system was that of classical biochemistry: to reproduce ATP-dependent protein breakdown in a cell-free system and then to fractionate such a system and to find the mode of action of its components.

An ATP-dependent proteolytic system from reticulocytes was first described by Etlinger and Goldberg[7], and then was analyzed by our biochemical fractionation–reconstitution studies. In this work, I was greatly helped by Aaron Ciechanover, who was then my graduate student. Substantial support and advice were provided by Irwin Rose, who hosted me in his laboratory in Fox Chase Cancer Center for a sabbatical year in 1977–1978 and many times afterwards. Initially, reticulocyte lysates were fractionated on DEAE-cellulose into two crude fractions: fraction 1, which was not adsorbed, and fraction 2, which contained all proteins adsorbed to the resin and eluted with high salt. The original aim of his fractionation had been to remove hemoglobin (present in fraction 1), but we found that fraction 2 lost most of ATP-dependent proteolytic activity. Activity could be restored by combining fractions 1 and 2. The active component in fraction 1 was a small protein that we purified

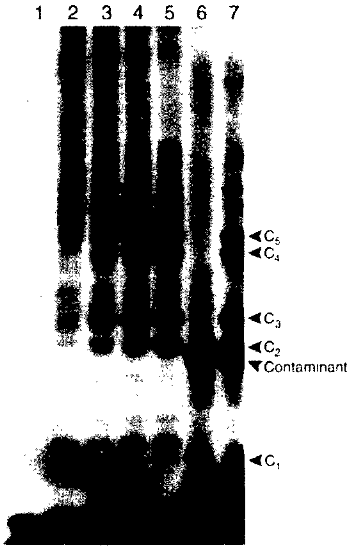


Fig. 1. Discovery of the ligation of ubiquitin to lysozyme, a substrate of the proteolytic system. Reaction products were separated by SDS–PAGE. Lane 1, incubation of ^{125}I-labeled ubiquitin with fraction 1 in the absence of ATP; ubiquitin remains free and migrates at the front. Lanes 2–5, incubation of ^{125}I-labeled ubiquitin with fraction 1 in the presence of ATP. Lane 2, ubiquitin becomes covalently linked to many high-molecular-weight derivatives, presumably endogenous protein substrates present in fraction 2. Lanes 3–5, several new labeled bands appear (C1–C5), which increase with increasing concentrations of lysozyme. Lanes 6 and 7, incubation of ^{125}I-labeled lysozyme with fraction 2 in the absence of ATP (lane 6) or with ATP and unlabeled ubiquitin (lane 7); bands C1–C5 contain the label of ^{125}I-labeled lysozyme and consist of increasing numbers of ubiquitin molecules ligated to lysozyme. Reproduced from ref. 13, with permission.

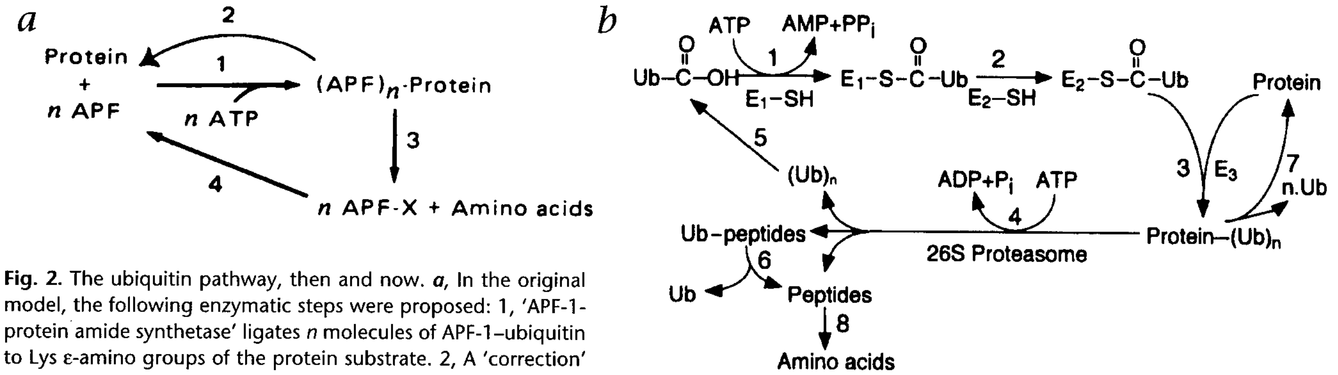


Fig. 2. The ubiquitin pathway, then and now. *a*, In the original model, the following enzymatic steps were proposed: 1, 'APF-1-protein amide synthetase' ligates *n* molecules of APF-1–ubiquitin to Lys ε-amino groups of the protein substrate. 2, A 'correction' amidase (isopeptidase) releases free protein and APF-1–ubiquitin from erroneous ligation products. 3, An endopeptidase (protease) specifically acts on proteins ligated to several molecules of APF-1 and cleaves peptide bonds with the liberation of APF-1 still linked to Lys or a Lys-containing peptide (APF-1-X). 4, Amidase (isopeptidase) cleaves the bond between APF-1 and the ε-amino group of Lys residues and thus liberates reusable APF-1–ubiquitin. Reproduced from ref. 13, with permission. *b*, Current information on the enzymatic reactions of the ubiquitin system. Steps 1, 2 and 3, accomplished by E1, E2 and E3, correspond to step 1 of the original model. Step 4, accomplished by the 26S proteasome, corresponds to step 3 of the original model. Steps 5, 6 and 7, accomplished by ubiquitin–carboxy-terminal hydrolases (isopeptidases), correspond to steps 2 and 4 of the original hypothesis. Reproduced from ref. 15, with permission.

by taking advantage of its remarkable stability to heat treatment[8]. It was first called APF-1, for ATP-dependent proteolysis factor 1. The identification of APF-1 as ubiquitin was later made by Wilkinson and co-workers[9], after our discovery of its ligation to proteins. Ubiquitin was first thought to be a thymic hormone, but subsequently was found to be present in many tissues and organisms, hence its name[10]. It was found to be conjugated to histone 2A (ref. 11), but its functions remained unknown. Although we did not know at that time that APF-1 was ubiquitin, I will use the term ubiquitin here to facilitate the discussion.

The purification of ubiquitin from fraction 1 was the key to the elucidation of the mode of its action in the proteolytic system. At first I thought that it could be an activator, or a regulatory subunit of a protease or other enzyme component of the system present in fraction 2. To examine this possibility, purified ubiquitin was radioiodinated and incubated with crude faction 2 in the presence or absence of ATP. There was substantial ATP-dependent binding of ^{125}I-labeled ubiquitin to high-molecular-weight proteins by gel filtration chromatography[12]. However, a covalent amide linkage was unexpectedly formed, as shown by the stability of the 'complex' to treatment with acid, alkali, hydroxylamine or boiling with SDS and mercaptoethanol[12]. Analysis of the reaction products by SDS–PAGE showed that ubiquitin was ligated to many high-molecular-weight proteins. Because crude fraction 2 from reticulocytes contains not only enzymes but also endogenous substrates of the proteolytic system, we began to suspect that ubiquitin might be linked to protein substrates, rather than to an enzyme. In support of this interpretation, we found that proteins that are good (although artificial) substrates for ATP-dependent proteolysis, such as lysozyme, form several conjugates with ubiquitin[13]. In the original experiment that convinced us that ubiquitin is ligated to the protein substrate, similar high-molecular-weight derivatives were formed when ^{125}I-labeled ubiquitin was incubated with unlabeled lysozyme (Fig. 1, lanes 3–5), and when ^{125}I-labeled lysozyme was incubated with unlabeled ubiquitin (Fig. 1, lane 7). Analysis of the ratio of radioactivity in ubiquitin and lysozyme indicated that the various derivatives consisted of increasing numbers of ubiquitin molecules linked to one molecule of lysozyme. On the basis of these findings, we proposed a model in 1980 (Fig. 2*a*) in which several molecules of APF-1–ubiquitin are linked to Lys ε-amino groups of the protein substrate by an 'APF-1-protein amide synthetase' (Fig. 2*a*, step 1). We proposed that proteins ligated to several ubiquitins were broken down by a specific protease that recognizes such conjugates (Fig. 2*a*, step 3). Thus, the protein would be broken down to free amino acids and to APF-1–ubiquitin still linked by isopeptide linkage to Lys or a small peptide (APF-1-X). Finally, free APF-1–ubiquitin is released for re-use by the action of a specific amidase/isopeptidase (Fig. 2*a*, step 4). Based on a suggestion by Ernie Rose, we added a hypothetical 'correcting' isopeptidase to this scheme, which would release free ubiquitin and substrate protein from products of erroneous ubiquitin–protein ligation (Fig. 2*a*, step 2). An isopeptidase that may have such correction function was described recently[14].

Comparison of the original model with our current knowledge of the reactions of the ubiquitin pathway[15] (Fig. 2*b*) shows that the original model was essentially correct, but much further detail provides explanation for the high selectivity of ubiquitin-mediated protein degradation. Thus, we have found that 'APF-1-protein amide synthetase' is actually composed of three types of enzymes: a ubiquitin-activating enzyme E1, a ubiquitin-carrier protein E2 and a ubiquitin-protein ligase E3 (ref. 16). Specific E3 enzymes recognize specific structural features in specific protein substrates, and thus account for substrate selectivity[17]. Proteins ligated to multi-ubiquitin chains are degraded by a 26S proteasome complex discovered by Rechsteiner and co-workers[18]. ATP is needed not only for the ubiquitin–protein ligation reaction, as originally proposed, but also for the action of the 26S proteasome[18]. Finally, free and reusable ubiquitin is released by the action of a large variety of ubiquitin C-terminal hydrolases[15].

As indicated before[19], the main lesson from our story is the continued importance of the use of biochemistry in modern biomedical research. Without biochemistry, it is doubtful whether an entirely new system could have been discovered. On the other hand, molecular genetics has been essential in discovering the many functions of this system in processes such as cell cycle control, signal transduction and the immune response[15].

The ubiquitin–proteolytic pathway: From obscurity to the patient bed

AARON CIECHANOVER

Traditionally, researchers in the field of proteolysis have tried to characterize and purify a single protease while pursuing the activity they are studying. I used a similar approach at the beginning of my studies as a graduate student with Avram Hershko, when we initiated our efforts to purify the non-lysosomal, ATP-dependent proteolytic activity that Etlinger and Goldberg[7] and we[20] had characterized earlier in reticulocyte lysate. An obvious first step during the purification of proteins from red blood cells is to remove hemoglobin, the main protein in the extract, by anion exchange chromatography. When we fractionated the lysate, we could not recover the proteolytic activity in either the flow-through material that contained hemoglobin (fraction I) or in the adsorbed, high-salt-desorbed material (fraction II). The proteolytic activity could be recovered, however, after the addition of both fractions I and II to the reaction mixture[8] (Table 1). This was a critical experiment, as it taught us that this complex system contained at least two (and if two, possibly more, as was later shown to be true) complementary factors, and not a single, 'traditional' protease that was also ATP-dependent. From then on, we used the power of 'classical' biochemistry and a 'complementing-add-and-subtract' approach to initially purify from fraction I ATP-dependent proteolytic factor-1 (APF-1; ref. 21). We called the protein APF-1, as it was the first factor we characterized. At the same time we started to identify additional complementing factors, and looked for a simple terminology to enable convenient communication in the laboratory.

APF-1 was later identified as ubiquitin[22], a known protein of previously unknown function. Keith Wilkinson and Arthur Haas were post-doctoral fellows with Irwin A. Rose at the Fox Chase Cancer Center in Philadelphia, where Avram Hershko spent a sabbatical in 1977–1978 and, later, his summers. I joined him during the summers of 1978–1981. Wilkinson and Haas were fascinated by the new type of post-translational modification by APF-1 and, along with Michael Urban, who worked in a neighboring laboratory studying histones, identified APF-1 as ubiquitin. It was known that ubiquitin generates a single modified adduct with histones H2A and H2B; however, the function of these adducts has unexpectedly remained obscure to these very days. We adopted the existing terminology, and APF-1 became ubiquitin. In parallel, we showed that multiple molecules of the protein are conjugated covalently to the target substrate and suggested that they serve as a degradation signal[12,13]. Initially we thought that each ubiquitin moiety attaches to a single internal Lys in the target molecule. Later Hershko[23] and then Varshavsky[24] and their colleagues demonstrated the formation of a poly-ubiquitin chain anchored to a single internal Lys residue. In the conjugation reaction, the C-terminal group of ubiquitin (Gly) generates a high-energy, isopeptide bond with an ε-amino group of an internal Lys residue of the substrate. Later we showed that the first conjugation event can also involve the N-terminus residue of the protein. We then went on to dissect the mechanism that underlies the activation of ubiquitin that must precede the generation of the high-energy bond with the target protein. Here, we made use of the known mechanism of amino-acid activation during protein synthesis, but mostly of the mechanism shown by Fritz Lipmann and his colleagues when they reconstituted a cell-free, non-ribosomal biosynthesis of the bacterial deca-peptide antibiotic gramicidin S (ref. 25). The similarities between the three mechanisms of activation of animo acids for protein and peptide synthesis and ubiquitin were strikingly similar. Dissection of the activation mechanism of ubiquitin paved the road to purification, by reaction-based, 'covalent' affinity chromatography over immobilized ubiquitin, of the three sequentially acting, conjugating enzymes: the ubiquitin-activating enzyme E1, the ubiquitin carrier protein E2 and the ubiquitin protein ligase E3, which accomplishes the last and most-essential step in the conjugation reaction, specific ligation of ubiquitin to the target protein[16,26]. In 1980, before the identification of APF-1 as ubiquitin and the dissection of the conjugation mechanism, we had already proposed a model[13] (Fig. 2a) according to which conjugation of multiple APF-1 molecules targets the substrate to degradation by an unknown, yet conjugate-specific, downstream protease. Intact APF-1 that can be re-used is recycled through the activity of isopeptidases. The protease, the 26S proteasome complex, was characterized and purified later by Martin Rechsteiner and colleagues[18]. In 1982, a more-detailed model was described, and the system began to be placed in its appropriate biological context[6]. In particular, its functions were analyzed and compared to those of the lysosomes involved mostly in the degradation of extracellular proteins taken up through pinocytosis and receptor-mediated endocytosis[6].

The physiological relevance of our initial findings remained unknown, as until that time (1981) we did all our studies in a cell-free reconstituted system, using secretory (and not intracellular; as these were available in large amounts and low cost) proteins as model substrates. The first evidence that the system is involved in degradation of proteins *in vivo* came from immunochemical analysis of ubiquitin adducts in cells[27]. Using antibodies raised against ubiquitin, we showed that, after incubation of cells in the presence of amino-acid analogs, the resulting abnormal proteins are short-lived. Their rapid degradation is accompanied by a transient, yet substantial increase in the level of ubiquitin adducts, strongly indicating that they serve as essential intermediates in the proteolytic process. Later, stronger and more-direct proof of the involvement of the ubiquitin system in the degradation cellular proteins came from the observation Alexander Varshavsky, Daniel Finely and I made that a cell-cycle arrest mutant that contains a thermolabile E1 enzyme is also defective in the degradation of short-lived abnormal proteins at the non-permissive temperature[28,29].

An important yet unresolved problem at that time involved the identification of the specific signaling motifs that target proteins for degradation. My initial entry into this fascinating area

Table 1 Anion exchange chromatographical resolution of reticulocyte lysate into unadsorbed (fraction I), and adsorbed, high-salt-desorbed (fraction II) complementing proteolytic activities.

Enzyme fraction used	Percent degradation of labeled globin (–) ATP	Percent degradation of labeled globin (+) ATP
Complete lysate	1.5	10.0
Fraction I (flow-through material)	0	0
Fraction II (high salt desorbed eluate)	1.5	2.7
Fraction I + Fraction II	1.6	10.6

Adapted with modifications from ref. 8.

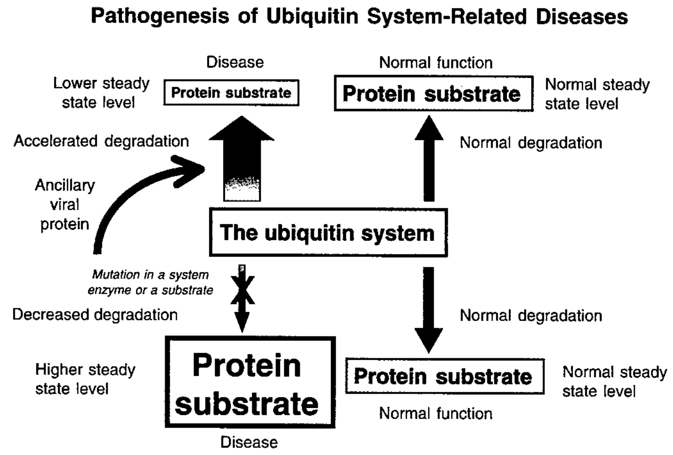


Fig. 3 Accelerated or decreased rates of ubiquitin-mediated proteolysis can underlie the pathogenesis of human diseases. Protein substrates are degraded each at a distinct and specific rate that may vary in different conditions (blue arrows), and maintain a steady-state level (green boxes) that enables them to function properly. Accelerated degradation (thick red arrow), as occurs in HPV E6-targeted degradation of p53, results in a low steady-state level of the target substrate (small beige box) and exposure of the cell to malignant transformation. Decreased degradation (narrow red arrow with a blue X) can occur when the signaling motif in the substrate is mutated (mutations in the phosphorylation sites in β-catenin in certain cases of malignant melanoma or colorectal carcinoma, or mutations in the NEDD4 E3 recognition motif of the kidney epithelial sodium channel in Liddle hypertension syndrome); or when E3 is mutated (as in Angelman syndrome, in which E6-AP is mutated). In all these examples, the excess accumulated substrate (large beige box) is 'toxic'.

was unanticipated. A large-scale purification of APF-1 carried out in Haifa in 1980 showed a substantial discrepancy between its high dry weight and low protein content, as determined by a Lowry assay. Hershko and I thought that the protein was probably a ribonucleoprotein complex, and the excess non-protein mass was due to the RNA component. Treatment of the APF-1 preparation with RNase A led to abrogation of its stimulatory activity towards bovine serum albumin (BSA), but not lysozyme. Not appreciating at that time the extent of the high specificity of the system toward its different substrates, we could not explain the 'selective' RNase effect, and did not pursue the study. The discrepancy between weight and protein measurement was resolved later with the finding that ubiquitin has a low content of the aromatic amino acids that are the basis of every known method for protein measurement. Obviously, ubiquitin is a pure protein and not a ribonucleoprotein complex. Later, our studies showed that nuclease added to the ubiquitin preparation destroyed $tRNA^{Arg}$ that was necessary, along with arginyl–tRNA protein transferase, to convert the N-terminal acidic (Asp) residue of BSA to Arg (refs. 30,31). Only the modified BSA and not the wild-type BSA can bind to the 'basic' N-terminal binding site of E3α, the ubiquitin ligase involved in recognition via the N-terminal residue (the N-end rule pathway ligase; ref. 15 and see below commentary by A. Varshavsky). Lysozyme, with a Lys at the N-terminal residue, does not undergo this post-translational modification and is recognized directly by E3α, and its degradation, therefore, is not sensitive to RNase. This finding became part of the more thorough and systematic mode of recognition identified through genetic tools by Alexander Varshavsky and his colleagues, and is known as the N-end-rule pathway[32]. Earlier, Hershko also noted the importance of an exposed N-terminal residue in targeting certain model proteins for degradation[33], but at that time, the mechanistic relevance of this finding was not apparent. Although most known substrates of the ubiquitin system are targeted through different recognition motifs[34], the N-end-rule pathway was the first well-defined signal of substrate recognition (see commentary by A. Varshavsky).

As for the myriad cellular substrates and regulatory functions of the ubiquitin system now known, it was only in the early 1990s that scientists started to unravel these secrets. The discovery of oncoproteins, tumor suppressors, transcriptional factors and cell-cycle regulators have shown that all these proteins are short-lived and their stability is tightly regulated. This previously unknown mode of regulation through destruction, which, unlike phosphorylation, is irreversible and may have evolved to secure directionality, has attracted many scientists to study the systems and signals that govern the stability (and hence the activity) of many different proteins[35–38]. Without even knowing the underlying mechanisms, Varshavsky, Finley and I predicted that the ubiquitin system may be involved in the regulation of the cell cycle[28]. The prediction was based on the observation that the ts85 mutant cell is defective in both E1 that inactivates the ubiquitin system and in transition along the S/G2 boundary of the cell cycle. This prediction has been corroborated by many studies demonstrating involvement of the ubiquitin system in programmed degradation of a broad array of cell cycle regulators.

Given the many processes and substrates involved, recent indications of the involvement of the system in the pathogenesis of many inherited as well as acquired diseases are not unexpected. Drug companies are trying to target the aberrations in the system that underlie these pathologies. The common denominator shared by all these diseases is a change in the steady-state level of a particular protein substrate or set of protein substrates. In general, the diseases belong to two classes, resulting from accelerated or decreased rates of degradation of different substrates (Fig. 3). The second class can be further divided into diseases due to mutations in enzymes of the system or to mutations in recognition motifs of substrates (Fig. 3). Although it is not possible to systematically review here all these diseases (reviewed recently in ref. 39), a few salient examples follow. An example of accelerated degradation involves the degradation of p53 induced by the human papillomavirus (HPV) E6 oncoprotein, which probably underlies the pathogenesis of human uterine cervical carcinoma, a very prevalent and severe malignant disease. E6 associates with p53 and targets it for rapid degradation mediated by the ubiquitin ligase E6-associated protein (E6-AP), which does not recognize the free tumor suppressor. Inactivation of the cellular DNA damage-control machinery exposes the cell to malignant transformation. Mutations in E6-AP lead to Angelman syndrome, an inherited disease associated with severe mental retardation and motor disorders. Here, accumulation of unidentified native substrate(s) of E6-AP (p53 is not the native substrate of the enzyme and is normally targeted by Mdm2) is probably toxic to the developing brain. In another example, mutations in the phosphorylation-targeting motif of the transcription factor β-catenin, or in adenomatosis polyposis coli (APC) that is part of the β-catenin degradation complex, lead to stabilization and accumulation of the protein, accompanied by its uncontrolled activity. These mutations may be involved in the pathogenesis of many forms of colorectal carcinomas and malignant melanomas. Viruses such as HPV have evolved different mechanism that enable them to evade the normal mode of

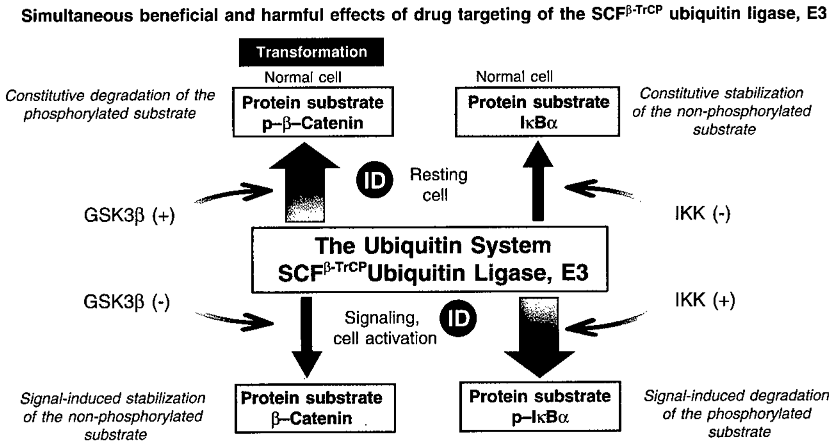

Simultaneous beneficial and harmful effects of drug targeting of the SCFβ-TrCP ubiquitin ligase, E3
Transformation
Normal cell
Normal cell
Constitutive degradation of the phosphorylated substrate
Protein substrate p–β–Catenin
Protein substrate IκBα
Constitutive stabilization of the non-phosphorylated substrate
ID
Resting cell
GSK3β (+)
IKK (-)
The Ubiquitin System SCFβ-TrCPUbiquitin Ligase, E3
GSK3β (-)
IKK (+)
Signaling, cell activation
ID
Signal-induced stabilization of the non-phosphorylated substrate
Protein substrate β–Catenin
Protein substrate p–IκBα
Signal-induced degradation of the phosphorylated substrate

Fig. 4 Drug targeting of an E3 enzyme can become a 'double-edged sword'. In the resting cell (top), IκBα (blue arrow and beige box, upper right) is not phosphorylated and degraded slowly, whereas β-catenin is phosphorylated constitutively by GSK3β and is degraded rapidly following ubiquitination by the $\mathrm{SCF}^{\beta\text{-TrCP}}$–E3–ubiquitin–ligase complex (heavy red arrow and light blue box, upper left). In the signaled cell (bottom), IκBα is phosphorylated by IκB kinase (IKK), rapidly ubiquitinated by the same SCF–TrCP complex and degraded (thick red arrow and light blue box, lower right). Also, in signaled cells, GSK3β is inhibited, and the non-phosphorylated β-catenin is stabilized, translocated into the nucleus and stimulates transcription (blue arrow and beige box, lower left). In general, IκB-containing cells are distinct from β-catenin-containing cells and so are the signals that activate the two pathways. However, the phosphorylated recognition motifs of the two proteins are similar and they seem to be targeted by the same TrCP ubiquitin ligase. An E3 inhibitor (inhibitory drug, ID) can lead to inhibition of degradation of IκB in stimulated cells, and consequently to suppression (beneficial effect) of NF-κB-induced inflammatory processes that may occur in autoimmune diseases, for example. At the same time, ID treatment will result in suppression of degradation of phosphorylated β-catenin in resting cells, with resultant accumulation of the transcription factor and possible subsequent malignant transformation (harmful effect).

activity of the ubiquitin system and allow them to continue their replication and propagation. Epstein Barr nuclear antigen 1 (EBNA-1) persists in healthy carriers for life, and its persistence contributes to some of the virus-related pathologies. Unlike all other Epstein Barr viral proteins, EBNA-1 cannot elicit a cytotoxic T lymphocyte (CTL) response. A long, C-terminal Gly-Ala repeat inhibits ubiquitin-mediated degradation and subsequent major histocompatibility complex (MHC) class I antigen presentation of EBNA-1. The human cytomegalovirus (CMV) encodes two endoplasmic reticulum (ER) resident proteins, US2 and US11, that bind to MHC class I molecules in the ER and escort them to the translocation machinery. After retrograde transport to the cytoplasm, they are ubiquitinated and degraded by the proteasome. Removal of the MHC molecules enables the virus to evade the immune system. A completely different case involves Liddle syndrome. In this disorder, a mutation in the recognition motif that targets the kidney epithelial sodium channel (ENaC). to ubiquitination by the Nedd4 E3 leads to accumulation of the channel, excessive reabsorption of sodium and water, with resulting severe hypertension.

Because of the central function of the ubiquitin system in many basic cellular processes, development of drugs that modulate the system may be difficult. Inhibition of enzymes common to the entire pathway, such as the proteasome, may affect many processes nonspecifically, although a narrow 'window' between beneficial effects and toxicity can be identified for a short-term treatment. An attractive possibility is the development of small molecules that inhibit specific E3 molecules. For example, specific phospho-peptide derivatives can inhibit the β-TrCP ubiquitin ligase, E3 complex (β-Transducin repeat-Containing Protein[40]). However, this approach can turn into a 'double-edged sword' (Fig. 4). Ideally, small molecules should be developed that bind to specific substrates or to their ancillary proteins, and thus inhibit a specific process. Peptide aptamers (small molecules/peptides that bind to active/association sites of proteins and inhibit their native interactions) that bind specifically to HPV E6 and probably prevent its association with p53, have been shown to induce apoptosis and reverse certain malignant characteristics in HPV-transformed cells, probably by interfering with p53 targeting[41]. Unfortunately, because of the rarity of proteins targeted by similar mechanisms, this approach may be currently limited to a small number of cases.

Discovering the functions and degrons of the ubiquitin system

ALEXANDER VARSHAVSKY

Through preparation, help from friends and a lot of luck I was able to leave the former Soviet Union in the fall of 1977, and ended up in Boston. A month later I was a faculty member of the Biology Department of the Massachusetts Institute of Technology (MIT), before I knew what exactly grants were (and before the colleagues who hired me became aware of that fact). In Moscow, I studied chromosome structure and regulation of gene expression, and looked forward to continuing this work.

There were few similarities between my earlier milieu and the astonishing new life. The libraries were one of them. They were just as quiet and pleasant in Cambridge as in Moscow, and a library at MIT soon became my second home. Reading there I came across a curious 1977 paper by Harris Busch, Ira Godknopf and their colleagues. They found a DNA-associated protein that had one C-terminus but two N-termini, an unprecedented structure. The short arm of that Y-shaped protein was joined, through its C-terminus, to an internal Lys of histone H2A. The short arm was soon identified, by Margaret Dayhoff, as ubiquitin, a 76-residue protein of unknown function that was described (as a free protein) by Gideon Goldstein and colleagues in 1975 (ref. 10).

I became interested in this first ubiquitin conjugate, UbH2A. Back in Russia, I had begun to develop a method for high-resolution analysis of nucleosomes. These DNA–protein complexes were subjected to electrophoresis in a low-ionic-strength polyacrylamide gel (a forerunner of the gel-shift assay), followed by second-dimension electrophoresis of either DNA or proteins. We located UbH2A in a subset of the nucleosomes, succeeded in separating these nucleosomes from those lacking UbH2A, and eventually showed that UbH2A-containing nucleosomes were enriched in transcribed genes and excluded from the inactive (heterochromatic) parts of the chromosomes[9].

Meanwhile, Avram Hershko, his graduate student Aaron Ciechanover and their colleagues at the Technion (Haifa, Israel) were studying ATP-dependent protein degradation in extracts from rabbit reticulocytes. In 1978–1980 they demonstrated that a small protein, which they called APF-1 (ATP-dependent proteolytic factor 1), was covalently conjugated to proteins that were about to be degraded in the extract. They suggested that a protein-linked APF-1 served as a signal for a downstream protease, and began the analysis of enzymology of APF-1 conjugation. In 1980, Keith Wilkinson, Michael Urban and Arthur Haas showed that APF-1 and ubiquitin were the same protein[9].

When I saw that 1980 paper, two seemingly independent realms, protein degradation and chromosomes, came together. I realized that we were dealing with a proteolytic system of immense complexity and exceptionally broad, still to be discovered, range of functions. I decided to find genetic approaches to this entire problem, because a system of such complexity was unlikely to be understood through biochemistry alone. In 1980, reverse-genetic techniques were about to become feasible with the yeast *Saccharomyces cerevisiae*, but were still a decade away from mammalian genetics. I continued to read, as widely as I could. On a fateful day at the end of 1980, I came across a paper by Yamada and colleagues that described a conditionally lethal, temperature-sensitive mouse cell line called ts85. The researchers showed that a specific nuclear protein disappeared from ts85 cells at increased temperatures, and suggested that this protein might be UbH2A. When I saw their data, I had to calm down to continue reading, because I knew that this protein was UbH2A. (In the preceding two years we had learned much about the electrophoretic properties of UbH2A.)

Daniel Finley had just joined my lab to study regulation of gene expression, but soon switched to ts85 cells. A few months into the project, Finley and I made the crucial observation that ubiquitin conjugation in an extract from ts85 cells was temperature-sensitive, in contrast to an extract from parental cells. Soon afterward, I invited Ciechanover, who came from the Hershko laboratory for a postdoctoral stint at another MIT lab, to join Finley and me in the continuing study of ts85 cells. He did, and we published two papers in 1984 that demonstrated two main results: that mouse ts85 cells have a temperature-sensitive, ubiquitin-activating (E1) enzyme, and that these cells stop degrading the bulk of their normally short-lived proteins at the nonpermissive temperature[28,29]. This was the first evidence that ubiquitin conjugation was required for protein degradation *in vivo*. These findings[28,29] also indicated that ubiquitin conjugation was essential for cell viability. In addition, ts85 cells tended to be arrested at the G2 phase of the cell cycle, and the synthesis of heat-shock proteins was strongly induced in these cells at the nonpermissive temperature, indicating that ubiquitin-dependent proteolysis is involved in the cell-cycle progression and stress response[28]. In 1983, Tim Hunt and colleagues discovered unusual proteins in rapidly dividing fertilized clam eggs. These proteins, which they called cyclins, were degraded at the exit from mitosis. We suggested in 1984 that cyclins were destroyed by the ubiquitin system[29], a hypothesis shown to be correct by Michael Glotzer, Andrew Murray and Marc Kirschner in 1991.

The ts85 results[28,29] left little doubt, among the optimists, about the importance of the ubiquitin system in cellular physiology. Unfortunately, these findings could not be deepened and made more rigorous, because of limitations of mammalian somatic cell genetics, which was still hampered at that time by the impossibility of altering genes at will. Therefore, in 1983 we began systematic analysis of the ubiquitin system in *S. cerevisiae* (Fig. 5). In 1984, Finley and Engin Özkaynak cloned the first ubiquitin gene, and found that it encoded a polyubiquitin precursor protein. By 1987, they showed that this gene, *UBI4*, was strongly induced by different stresses. Moreover, deletion of *UBI4* resulted in cells that were hypersensitive to every noxious treatment we tried, including heat and oxidative stress[42]. These results validated and extended an inference from the 1984 findings with ts85 cells, thereby establishing one broad and essential function for the ubiquitin system.

In a parallel 1987 study, Stefan Jentsch and John McGrath isolated ubiquitin-conjugating (E2) enzymes from *S. cerevisiae*. One evening, a phone call from an excited Stefan Jentsch marked the discovery of yet another function of the ubiquitin system: a partially sequenced yeast E2 enzyme was found to be RAD6, a protein known to yeast geneticists for years as an essential component of DNA repair pathways[43]. RAD6 was the first enzyme of the ubiquitin system that was shown to mediate a specific physiological function. The sequence of RAD6 was weakly similar

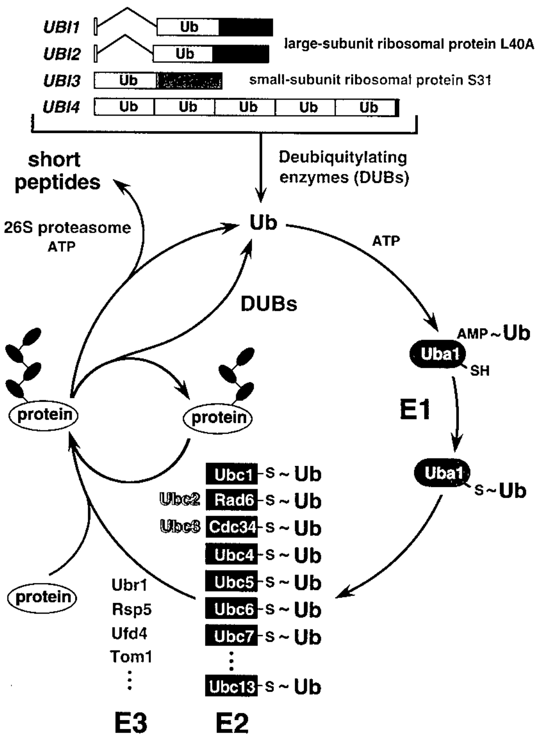


Fig. 5 The ubiquitin system of *S. cerevisiae*[51]. The yeast ubiquitin genes, two of which (*UBI1* and *UBI2*) contain introns, encode fusion proteins of ubiquitin (yellow rectangles) to itself (*UBI4*) or to one of the two specific ribosomal proteins (*UBI1–UBI3*) (red and blue rectangles). These fusion proteins are cleaved by deubiquitinating enzymes, yielding mature ubiquitin. ~, Thioester bonds between ubiquitin and the active-site Cys residues of ubiquitin-specific enzymes. The conjugation of ubiquitin to other proteins involves a preliminary ATP-dependent step, in which the last residue of ubiquitin (Gly76) is joined, through a thioester bond, to a Cys residue in the ubiquitin-activating (E1) enzyme encoded by *UBA1*. The activated ubiquitin is transferred to a Cys residue in one of at least 13 distinct ubiquitin-conjugating (E2) enzymes encoded by the *UBC* family genes, and from there to a Lys residue of an ultimate acceptor protein (yellow oval). This last step and the formation of a multi-ubiquitin chain (black ovals) require participation of another component, called E3 (the names of some of the yeast E3 proteins are included). A targeted, ubiquitinated protein substrate is processively degraded to short peptides by the ATP-dependent 26S proteasome.

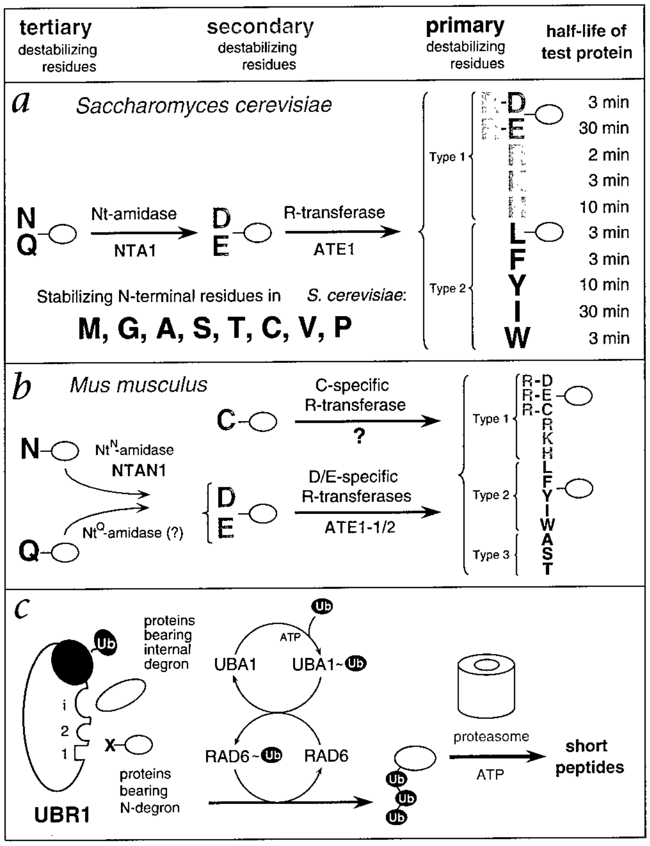


Fig. 6 The N-end rule pathway. Notations in the yeast (*a*) and mouse (*b*) pathways show type 1 (purple) and type 2 (red) primary, secondary (light blue) and tertiary (green) destabilizing N-terminal residues; yellow ovals indicate the rest of a protein substrate. ***a***, The *in vivo* half-lives of X-βgals, β-galactosidase-based test proteins in *S. cerevisiae*[48] (right). X-βgal proteins bearing stabilizing N-terminal residues (black) are metabolically stable ($t_{1/2}$, more than 20 h). The tertiary destabilizing residues N (Asn) and Q (Gln) are converted into secondary destabilizing residues D (Asp) and E (Glu) by N-terminal amidohydrolase (Nt-amidase), encoded by *NTA1*. D and E are conjugated to R (Arg), one of the primary destabilizing residues, by Arg–RNA protein transferase (R-transferase), encoded by *ATE1*. ***b***, In the mammalian N-end rule pathway, the deamidation step is mediated by two distinct enzymes, Nt^N-amidase and Nt^Q-amidase, specific for N-terminal Asn and Gln residues, respectively [54]. In vertebrates, the set of secondary destabilizing residues contains not only Asp and Glu but also Cys (C), which is a stabilizing residue in yeast[55]. In mammals but not in yeast, Ala (A), Ser (S) and Thr (T) are primary (type 3) destabilizing residues[48]. ***c***, *S. cerevisiae* UBR1 has two binding sites for the primary destabilizing N-terminal residues of either proteins or short peptides. The type 1 site is specific for basic N-terminal residues Arg, Lys and His. The type 2 site is specific for bulky hydrophobic N-terminal residues Phe, Leu, Trp, Tyr and Ile. UBR1 contains yet another substrate-binding site (i), which targets proteins bearing internal (non-N-terminal) degrons. In yeast, these proteins include the CUP9 repressor[57]. A complex of UBR1 and the ubiquitin-conjugating (E2) enzyme RAD6 produces a substrate-linked multi-ubiquitin chain[48].

to that of CDC34, an essential cell-cycle regulator defined genetically by Leland Hartwell. In 1988, a collaboration between Breck Byer's and my laboratories demonstrated that CDC34 was also a ubiquitin-conjugating enzyme[44] (Fig. 5). This result transformed a hint from our ts85 work into a definitive demonstration of the involvement of the ubiquitin system in cell-cycle control.

In 1989, Finley and Bonnie Bartel discovered that ubiquitin genes other than *UBI4* (the polyubiquitin gene) were also quite unusual: *UBI1–UBI3* encoded fusions of ubiquitin to one protein of the large ribosomal subunit and one protein of the small ribosomal subunit, an arrangement conserved from yeast to humans[45] (Fig. 5). Kenneth Redman and Martin Rechsteiner independently identified these non-ubiquitin extensions as ribosomal proteins. The transient presence of ubiquitin in front of a ribosomal protein moiety (ubiquitin was rapidly cleaved off by deubiquitinating enzymes) was found to be essential for efficient biogenesis of the ribosomes[45]. Ubiquitin acts, in these settings, not as a degradation signal but as a molecular chaperone. The fusion-imposed 1:1 molar ratio of free ubiquitin to a free ribosomal protein (Fig. 5) sets an upper limit for the number of newly produced ribosomes relative to the number of newly formed ubiquitin molecules. This tight link, through DNA-encoded fusions of ubiquitin and ribosomal proteins, is one of the few understood regulatory interactions between protein synthesis and protein degradation.

The enormous expansion of the ubiquitin field in the last decade stemmed mainly from these functional insights of the 1980s, which demonstrated both the involvement of ubiquitin conjugation in important biological processes and the striking diversity of these processes, from the cell cycle[28, 29, 44] to DNA repair[43], ribosome biogenesis[45] and stress responses[42]. Many more functions have been added to this list since 1990.

How are proteins recognized as substrates for ubiquitin conjugation? The first solution to this problem was produced in 1986, when Andreas Bachmair and Finley discovered the first degradation signals in short-lived proteins[46]. We constructed ubiquitin fusion proteins in which ubiquitin was followed by a reporter moiety such as *Escherichia coli* β-galactosidase, and expressed them in *S. cerevisiae*. The first advance took place when we learned that the ubiquitin moiety of these fusion proteins was rapidly removed by deubiquitinating enzymes regardless of the identity of the residue at the C-terminal side of the cleavage site, with Pro being the sole exception. Thus was born the ubiquitin fusion technique, which made it possible to place, *in vivo*, any desired residue (except Pro) at the N-terminus of a protein of interest[46]. The presence of Met at the N-termini of nascent proteins and the substrate specificity of cytosolic Met aminopeptidases did not allow this level of experimental freedom before the discovery of the ubiquitin fusion technique[47].

Using this method, Bachmair and Finley discovered that the *in vivo* half-life of a test protein was strongly dependent on the identity of its N-terminal residue, a simple relation called the N-end rule[46]. The underlying ubiquitin-dependent pathway, called the N-end-rule pathway (Fig. 6), was later found to be present in all eukaryotes, from fungi to plants and mammals, and even in prokaryotes, which lack ubiquitin[48]. Yet another degradation signal identified in 1986 was the N-terminal ubiquitin moiety of a fusion protein under conditions that precluded its removal by deubiquitinating enzymes[46]. This signal is targeted by a distinct pathway of the ubiquitin system[49].

A family of signals, called N-degrons, that give rise to the N-end rule is still the best-understood set of degradation signals. An N-degron consists of a substrate's destabilizing N-terminal residue and an internal Lys residue, the latter being the site of ubiquitin attachment[48,50]. The E2–E3 ubiquitin ligase (Fig. 6*c*) binds to the substrate's N-terminal residue and forms a multi-ubiquitin chain linked to a substrate's Lys residue, the selection of which is often the result of stochastic choice among several sterically suitable Lys residues[48]. This bi-partite organization is also characteristic of subsequently identified degradation signals

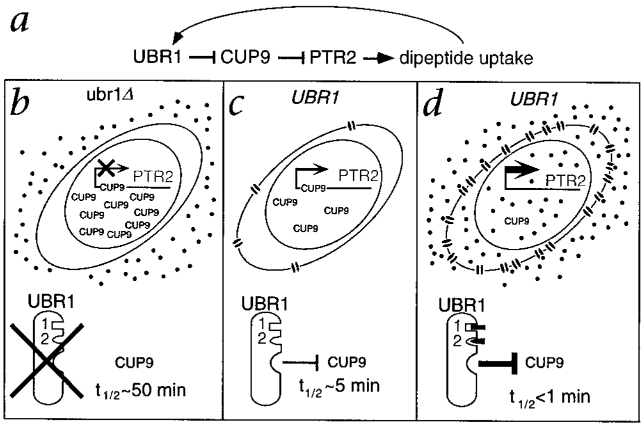


Fig. 7 Ubiquitin-dependent activation of peptide import in *S. cerevisiae*[20]. ***a,*** Genetic diagram of the peptide transport circuit. ***b,*** UBR1 is required for di-peptide uptake. In the absence of UBR1 (*ubr1Δ*), the transcriptional repressor CUP9 is long-lived, accumulates to high levels and extinguishes the expression of peptide transporter encoded by *PTR2*. The *ubr1Δ* cells cannot import di-peptides (red dots). ***c,*** In a *UBR1* cell growing in the absence of extracellular di-peptides, UBR1 targets CUP9 for degradation ($t_{1/2}$, about 5 min), resulting in a lower steady-state concentration of CUP9 and weak but substantial expression of the PTR2 transporter (blue double ovals). ***d,*** In *UBR1* cells growing in the presence of extracellular di-peptides, some of which bear destabilizing N-terminal residues, the imported di-peptides bind to the basic (type 1; red rectangle) or hydrophobic (type 2; green wedge) residue-binding sites of UBR1. Binding of either type of di-peptide to UBR1 allosterically increases the rate of UBR1-mediated degradation of CUP9. The resulting decrease of the half-life of CUP9 from about 5 min to less than 1 min leads to a further decrease in CUP9 levels, and consequently to a strong induction of the PTR2 transporter[57].

in cyclins, transcription factors and other short-lived proteins. One unique feature of N-degrons is that substrates bearing certain destabilizing N-terminal residues are chemically modified *in vivo*, through their enzymatic deamidation or arginylation, before a substrate can be bound by the ubiquitin ligase[48,51] (Fig. 6).

Having been the first ubiquitin-dependent pathway to be defined through molecular genetic methods, the N-end rule pathway (Fig. 6) was also the setting in which several essential insights relevant to the entire ubiquitin system were first made, including the discovery of specific multi-ubiquitin chains and their function in proteolysis[24]. In 1985, Hershko and Heller suggested, on the basis of chemical modification data, that some ubiquitin moieties in multi-ubiquitinated proteins might be linked together in a chain. In 1989, Vincent Chau and colleagues in my laboratory demonstrated the existence of protein-linked multi-ubiquitin chains, found them to have unique topology (ubiquitin–ubiquitin bonds through Lys48 of ubiquitin) and showed that these chains were required for degradation of test proteins[24]. We proposed that the main function of a substrate-linked multi-ubiquitin chain is to bind the substrate to the proteasome[24]. The complexity and multiplicity of ways in which a substrate is delivered to the proteasome is demonstrated by the recent discovery that ubiquitin ligases themselves physically interact with specific subunits of the 26S proteasome[52].

Subunit selectivity of protein degradation was yet another fundamental feature of the ubiquitin system that was first discovered in the N-end rule pathway. Erica Johnson and David Gonda demonstrated in 1990 that this pathway can eliminate one subunit of an oligomeric protein selectively, leaving intact the other subunits of the same protein molecule[53]. It is specifically the subunit conjugated to a multi-ubiquitin chain that gets destroyed. Subunit selectivity of proteolysis underlies large differences in the *in vivo* half-lives of subunits in oligomeric proteins. This essential feature of the ubiquitin system is both powerful and flexible, in that it allows protein degradation to be wielded as an instrument of either positive or negative control. Among many examples are activation of transcription factor NF-κB through degradation of its inhibitory ligand IF-κB, and inactivation of cyclin-dependent kinase activity through degradation of a regulatory cyclin subunit.

The emerging functions of the N-end rule pathway have been described[54,55]. Among these functions, the best understood is the essential role of this pathway in a positive feedback circuit that regulates the import of peptides in *S. cerevisiae*[56,57] (Fig. 7). Imported peptides bearing destabilizing N-terminal residues bind to the recognition sites for N-end rule substrates in UBR1, the pathway's E3 enzyme. This binding allosterically activates yet another substrate-binding site of UBR1, leading to accelerated degradation of the transcriptional repressor CUP9. The resulting derepression of expression of the peptide transporter PTR2 greatly increases the cell's capacity to import peptides[57]. This circuit (Fig. 7) is the first example of small compounds being natural allosteric regulators of the ubiquitin system.

A backward glance: It's Moscow, and the year is 1968. The author, a chemistry undergraduate both cocky and insecure, is listening to a leading Russian biochemist, a man in his forties whose education and entire life were warped by the combined cruelties of Stalinism and the Lysenko-led destruction of Russian genetics. The great man was telling me something he considered self-evident: "Ah, Alex, don't waste your time on genetics. It's all ancient Greece, beautiful in a strange way, but next to useless. They keep tormenting fruit flies, but it's us biochemists who will produce the understanding that really matters." Having spent a day reading genetic papers, I sensed that he could not be right, that genetics was essential too. Over the next three decades, the dynamic interaction of genetics and biochemistry kept yielding insights that could not be produced by biochemistry or genetics alone. These advances, many of them technical in nature, have transformed biology, and are beginning to be felt in medicine.

The early history of the ubiquitin field recapitulates, in a microcosm, the essential interaction between biochemistry and genetics that underlies the phenomenon of modern biology. Biochemical studies by Hershko, Ciechanover and their colleagues revealed a mechanistically unexpected, most curious but functionally obscure pathway of protein degradation. Molecular genetic (as well as biochemical) work proved necessary for discovering the first physiological functions of ubiquitin-dependent proteolysis and the first degradation signals in short-lived proteins. Methods and approaches developed in this work, including the ubiquitin fusion technique[47], continue to be of use in the ubiquitin field and beyond.

The vast expansion of ubiquitin studies over the last decade, with hundreds of laboratories around the world working on the ubiquitin system and its legion of biological functions, is a sight to behold. The fundamental understanding of this system, and recent insights into its roles in health and disease will have a profound influence on the realm of therapeutic drugs. The reason is not just the obvious one—promising drug targets among the ubiquitin system's components and substrates—but also the possibility of developing drugs that could direct this system to destroy (and thereby to inhibit functionally) any protein target.

1. Schoenheimer, R. *The Dynamic State of Body Constituents* (Harvard University Press, Cambridge, Massachusetts, 1942).
2. Schimke, R.T. & Doyle, D. Control of enzyme levels in animal tissues. *Annu. Rev. Biochem.* **39**, 929–979 (1971).
3. Haider, M. & Segal, H.L. Some characteristics of the alanine aminotransferase- and arginase-inactivating system of lysosomes. *Arch. Biochem. Biophys.* **148**, 228–237 (1972).
4. Hershko, A. & Tomkins, G.M. Studies on the degradation of tyrosine aminotransferase in hepatoma cells in culture. Influence of the composition of the medium and adenosine triphosphate dependence. *J. Biol. Chem.* **246**, 710–714 (1971).
5. Simpson, M.V. The release of labeled amino acids from proteins in liver slices. *J. Biol. Chem.* **201**, 143–154 (1953).
6. Hershko, A. & Ciechanover, A. Mechanisms of intracellular protein breakdown. *Annu. Rev. Biochem.* **51**, 335–364 (1982).
7. Etlinger, J.D. & Goldberg, A.L. A soluble ATP-dependent proteolytic system responsible for the degradation of abnormal proteins in reticulocytes. *Proc. Natl. Acad. Sci. USA.* **74**, 54–58 (1977).
8. Ciechanover, A., Hod, Y. & Hershko, A. a heat-stable polypeptide component of an ATP-dependent proteolytic system from reticulocytes. *Biochem. Biophys. Res. Commun.* **81**, 1100–1105 (1978).
9. Wilkinson, K.D., Urban, M.K. & Haas, A.L. Ubiquitin is the ATP-dependent proteolysis factor of rabbit reticulocytes. *J. Biol. Chem.* **255**, 7529–7532 (1980).
10. Goldstein, G. *et al.* Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. *Proc. Natl. Acad. Sci. USA.* **72**, 11–15 (1975).
11. Goldknopf, I.L. & Busch, H. Isopeptide linkage between nonhistone and histone A polypeptides of chromosomal conjugate protein A24. *Proc. Natl. Acad. Sci. USA.* **74**, 864–868 (1977).
12. Ciechanover, A., Heller, H., Elias, S., Haas, A. L. & Hershko, A. ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. *Proc. Natl. Acad. Sci. USA.* **77**, 1365–1368 (1980).
13. Hershko, A., Ciechanover, A, Heller, H., Haas, A. L. & Rose, I. A. Proposed role of ATP in protein breakdown: conjugation of proteins with multiple chains of the polypeptide of ATP-dependent proteolysis. *Proc. Natl. Acad. Sci. USA.* **77**, 1783–1786 (1980).
14. Lam, Y.A., Xu, W., DeMartino, G.N. & Cohen, R.E. Editing of ubiquitin conjugates by an isopeptidase of the 26S proteasome. *Nature* **385**, 737–740 (1997).
15. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425–479 (1998).
16. Hershko, A., Heller, H., Elias, S. & Ciechanover, A. Components of ubiquitin-protein ligase system: resolution, affinity purification and role in protein breakdown. *J. Biol. Chem.* **258**, 8206–8214 (1983).
17. Hershko, A., Heller, A., Eytan, E. & Reiss, Y. The protein binding site of the ubiquitin-protein ligase system. *J. Biol. Chem.* **261**, 11992–11999 (1986).
18. Hough, R., Pratt, G. & Rechsteiner, M. Ubiquitin-lysozyme conjugates. Identification and characterization of an ATP-dependent protease from rabbit reticulocyte lysates. *J. Biol. Chem.* **261**, 2400–2408 (1986).
19. Hershko, A. Lessons from the discovery of the ubiquitin system. *Trends Biochem. Sci.* **21**, 445–449 (1996).
20. Hershko, A., Heller, H., Ganoth, D. & Ciechanover, A. in *Protein Turnover and Lysosome Function* (eds. Segal, H.L. & Doyle, D.J.) 149–169 (Academic Press, New York, 1978).
21. Ciechanover, A., Elias, S., Heller, H., Ferber, S. & Hershko, A. Characterization of the heat-stable polypeptide of the ATP-dependent proteolytic system from reticulocytes. *J. Biol. Chem.* **255**, 7525–7528 (1980).
22. Wilkinson, K.D., Urban, M.K. & Haas, A.L. Ubiquitin is the ATP-dependent proteolysis factor I of rabbit reticulocytes. *J. Biol. Chem.* **255**, 7529–7532 (1980).
23. Hershko, A. & Heller, H. Occurrence of a polyubiquitin structure in ubiquitin-protein conjugates. *Biochem. Biophys. Res. Common.* **128**, 1079–1086 (1985).
24. Chau, V. *et al.* A multiubiquitin chain is confined to specific Lysine in a targeted short-lived protein. *Science* **243**, 1576–1583 (1989).
25. Lipmann, F, Gevers, W., Kleinkauf, H. & Roskoski, R.Jr. Polypeptide synthesis on protein templates: The enzymatic synthesis of gramicidin S and tyrocidine. *Adv. Enzymol. Relat. Areas Mol. Biol.* **35**, 1–34 (1971).
26. Ciechanover, A., Elias, S., Heller, H. & Hershko, A. "Covalent affinity" purification of ubiquitin activating enzyme. *J. Biol. Chem.* **257**, 2537–2542 (1982).
27. Hershko, A., Eytan, E., Ciechanover, A. & Haas, A.L. Immunochemical analysis of the turnover of ubiquitin-protein conjugates in intact cells: Relationship to the breakdown of abnormal proteins. *J. Biol. Chem.* **257**, 13964–13970 (1982).

28 Finley, D., Ciechanover, A. & Varshavsky, A.. Thermolability of ubiquitin-activating enzyme from the mammalian cell cycle mutant ts85. *Cell* **37**, 43–55 (1984).

29. Ciechanover, A., Finley D. & Varshavsky, A. Ubiquitin dependence of selective protein degradation demonstrated in the mammalian cell cycle mutant ts85. *Cell* **37**, 57–66 (1984).
30. Ferber, S. & Ciechanover, A. Transfer RNA is required for conjugation of ubiquitin to selective substrates of the ubiquitin- and ATP-dependent proteolytic system. *J. Biol. Chem.* **261**, 3128–3134 (1986).
31. Ferber, S. & Ciechanover, A. Role of arginine-tRNA in protein degradation by the ubiquitin pathway. *Nature* **326**, 808–811 (1987).
32. Varshavsky, A. The N-end rule pathway of protein degradation. *Genes Cells* **2**, 13–28 (1997).
33. Hershko, A., Heller, H., Eytan, E., Kaklij, G. & Rose, I.A. Role of α-amino group of protein in ubiquitin-mediated protein breakdown. *Proc. Natl. Acad. Sci. USA* **81**, 7021–7025 (1984).
34. Mayer, A. Siegel, N.R., Schwartz, A.L. & Ciechanover, A. Degradation of proteins with acetylated amino termini by the ubiquitin system. *Science* **244**, 1480–1483 (1989).
35. Scheffner, M., Werness, B.A., Huibregtse, J.M., Levine, A.J. & Howley, P.M. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* **63**, 1129–1136 (1990).
36. Glotzer, M., Murray, A.W. & Kirschner M.W. Cyclin is degraded by the ubiquitin pathway. *Nature* **349**, 132–138 (1991).
37. Hershko, A., Ganoth, D., Pehrson, J., Palazzo, R.E., & Cohen, L.H.. Methylated ubiquitin inhibits cyclin degradation in clam embryo extracts. *J. Biol. Chem.* **266**, 16376–16379 (1991).
38. Ciechanover, A. *et al.* Degradation of nuclear oncoproteins by the ubiquitin system *in vitro. Proc. Natl. Acad. Sci. USA* **88**, 139–143 (1991).
39. Ciechanover, A., Orian, A. & Schwartz, A.L.. Ubiquitin-mediated proteolysis: Biological regulation via destruction. *BioEssays* **22**, 442–451 (2000).
40. Yaron, A. *et al.* Inhibition of NF-κB cellular function via specific targeting of the IκBα-ubiquitin ligase. *EMBO J.* **16**, 6486–6494 (1997).
41. Butz, K., Denk, C., Ullmann, A., Scheffner, M. & Hoppe-Seyler, F. Induction of apoptosis in human papillomavirus positive cancer cells by peptide aptamers targeting the viral E6 oncoprotein. *Proc. Natl. Acad. Sci. USA* **97**, 6693–6697 (2000).
42. Finley, D., Özkaynak, E. & Varshavsky, A. The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation, and other stresses. *Cell* **48**, 1035–1046 (1987).
43. Jentsch, S., McGrath, J.P. & Varshavsky, A. The yeast DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme. *Nature* **329**, 131–134 (1987).
44. Goebl, M.G. *et al.* The yeast cell cycle gene *CDC34* encodes a ubiquitin-conjugating enzyme. *Science* **241**, 1331–1335 (1988).
45. Finley, D., Bartel, B. & Varshavsky, A. The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. *Nature* **338**, 394–401 (1989).
46. Bachmair, A., Finley, D. & Varshavsky, A. *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179–186 (1986).
47. Varshavsky, A. Ubiquitin fusion technique and its descendants. *Meth. Enzymol.* **327**, 578–593 (2000).
48. Varshavsky, A. The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. USA* **93**, 12142–12149 (1996).
49. Johnson, E. S., Ma, P. C., Ota, I. M. & Varshavsky, A. A proteolytic pathway that recognizes ubiquitin as a degradation signal. *J. Biol. Chem.* **270**, 17442–17456 (1995).
50. Suzuki, T. & Varshavsky, A. Degradation signals in the lysine-asparagine sequence space. *EMBO J.* **18**, 6017–6026 (1999).
51. Varshavsky, A. The ubiquitin system. *Trends Biochem. Sci.* **22**, 383–387 (1997).
52. Xie, Y. & Varshavsky, A. Physical association of ubiquitin ligases and the 26S proteasome. *Proc. Natl. Acad. Sci. USA* **97**, 2497–2502 (2000).
53. Johnson, E.S., Gonda, D.K. & Varshavsky, A. Cis-trans recognition and subunit-specific degradation of short-lived proteins. *Nature* **346**, 287–291 (1990).
54. Kwon, Y.T. *et al.* Altered activity, social behavior, and spatial memory in mice lacking the NTAN1p amidase and the asparagine branch of the N-end rule pathway. *Mol. Cell. Biol.* **20**, 4135–4148 (2000).
55. Davydov, I.V. & Varshavsky, A. RGS4 is arginylated and degraded by the N-end rule pathway in vitro. *J. Biol. Chem.* **275**, 22931–22941 (2000).
56. Byrd, C., Turner, G.C. & Varshavsky, A. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. *EMBO J.* **17**, 269–277 (1998).
57. Turner, G., Du, F. & Varshavsky, A. Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway. *Nature* **405**, 579–582 (2000).

Avram Hershko
Unit of Biochemistry, Technion-Israel Institute of Technology
Faculty of Medicine, P.O. Box 9649, Haifa, 31096, Israel

Aaron Ciechanover
Unit of Biochemistry, Technion-Israel Institute of Technology
Faculty of Medicine, P.O. Box 9649, Haifa, 31096, Israel

Alexander Varshavsky
Smits Professor of Cell Biology, Division of Biology
California Institute of Technology
1200 East California Blvd., Pasadena, CA 91125

Clinical Medical Research Award

Hepatitis C Virus and eliminating post-transfusion hepatitis

HARVEY J. ALTER & MICHAEL HOUGHTON

You'll wonder where the yellow went: A 30-year perspective on the near-eradication of post-transfusion hepatitis

HARVEY J. ALTER

The story I will relate here traces the near-total eradication of transfusion-associated hepatitis over the course of three decades. I am perhaps the thread that links these events, but the story is a fabric woven by many collaborators who played essential parts and by the conducive environment of the National Institutes of Health (NIH) intramural program that has nurtured these clinical investigations. The story and my research career began in the early 1960s when, as an NIH Clinical Associate, I investigated the non-cellular causes of febrile transfusion reactions by screening multiply transfused patients for antibodies against serum proteins using agar gel diffusion. Ouchterlony plates were piled high on my lab bench, the way unread manuscripts are today. One day Richard Aster told me that he had heard an interesting lecture by Baruch Blumberg, who was using similar methodology to investigate protein polymorphisms. I visited Blumberg and began a collaboration that within a year uncovered an unusual precipitin line resulting from the reaction between sera from a patient with hemophilia and an Australian Aboriginal person (Fig.1). The line was unusual in that it stained only faintly with lipid dyes, in contrast to the lipoprotein polymorphisms that were then being studied. Because it stained red with the azocarmine counter-stain, we initially called this the red antigen, then debated calling it the Bethesda antigen and ultimately called it the Australia antigen, based on the evolving nomenclature for new hemoglobin discoveries. Subsequent investigations showed the prevalence of Australia antigen to be only 0.1% in the donor population, but very high (10%) in patients with leukemia. The first publication on the Australia antigen[1] cited this association with leukemia in the title. We considered that the antigen might be a component of the long-postulated leukemia virus. In retrospect, the antigen merely reflected the high transfusion exposure and the immunocompromised status of these patients. My initial first-author publication was the biophysical characterization of the Australia antigen[2].

In 1964, I left the NIH to complete my training in internal medicine and hematology, and Blumberg moved to the Institute for Cancer Research in Philadelphia, where he continued to pursue the importance of the Australia antigen. Both serendipity and good science led, by 1968, to the linkage of this antigen to viral hepatitis, a link that transformed the study of hepatitis, protected the blood supply, led to a hepatitis B vaccine and culminated in the Nobel prize.

I returned to the NIH in 1969 to investigate the causes and prevention of post-transfusion hepatitis and to pursue clinical investigations of hepatitis B and its associated antigens. I had no premonition that this would be a lifetime endeavor. My first function was to continue and expand on prospective studies of post-transfusion hepatitis initiated by John Walsh, Bob Purcell, Paul Holland and Paul Schmidt. Walsh and colleagues had already demonstrated the inordinately high hepatitis risk of blood transfusion and, particularly, the risk of paid-donor blood. In 1970, Holland, Schmidt and I, in a still-memorable meeting that would later influence national blood policy, decided that the continued use of paid donors could not be tolerated and also concluded that we should introduce donor screening for what was by then called the hepatitis-associated antigen. I then simultaneously did a retrospective analysis that demonstrated the value of hepatitis B antigen testing and initiated a new prospective study to assess the effect of this dual change in the donor supply. The result was substantial: hepatitis incidence among patients undergoing open-heart surgery plummeted from 33% to 9.7% (ref. 3) (Fig. 2). We calculated that the main determinant of this reduction of about 70% was the exclusion of paid donors. Indeed, retrospective testing for hepatitis B virus (HBV) markers showed that only 20% of the hepatitis found before antigen screening was related to HBV. The recognition of transfusion-associated non-B hepatitis therefore evolved. Improved hepatitis B antigen assays brought hepatitis B transmission to near zero by 1977. In 1973, Steve Feinstone, who had the unenviable task of sifting through stool specimens in the bowels of NIH building 7, used immune electron microscopy to discover the hepatitis A virus, in collaboration with Al Kapikian and Bob Purcell[4]. We immediately delved into our repository of non-B hepatitis cases and were surprised to find that not a single case was due to hepatitis A virus[5]. In a less-than-brilliant foray into nomenclature, we designated these cases non-A, non-B (NANB) hepatitis. Bob Purcell, in particular, felt that we should not call the agent the hepatitis C virus until we had proved transmissibility and until we established the number of agents that might be involved. In our optimism, we did not suspect that the designation non-A, non-B would persist for 15 years before its specific etiology could be defined.

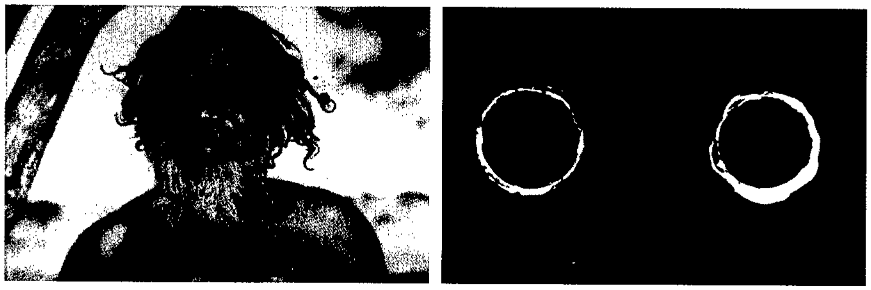


Fig.1 An Australian Aborigine (left) and the precipitin line formed between the aboriginal serum and that of a multiply transfused patient with hemophilia (right). The precipitin failed to stain for lipid, but stained red with the azocarmine counter-stain for protein.

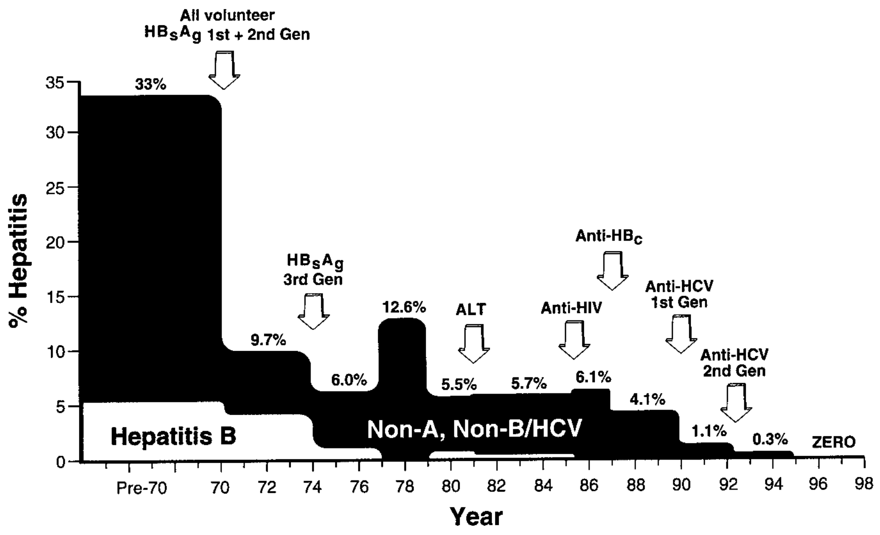


Fig. 2 The decreasing incidence of transfusion-associated hepatitis in blood recipients monitored prospectively. Incidence, traced from 1969 to 1998, demonstrates a decrease in risk from 33% to nearly zero. Arrows, main interventions in donor screening and selection that effected this change.

In 1975, as the prospective studies continued, my attention was directed at proving NANB hepatitis transmission in the chimpanzee model. Earlier attempts at chimpanzee transmission had failed, but I reasoned that we could use the highly pedigreed samples from our prospective studies and use inocula in volumes equivalent to blood transfusion in humans. We achieved success in our first attempt, as five of five chimpanzees developed increases in alanine aminotransferase (ALT) at appropriate intervals after inoculation[6]. We were later able to use this same approach and accomplish the first experimental animal transmission of human immunodeficiency virus[7]. In the absence of a tissue culture system, an observed particle or a serologic assay, the availability of an animal model was essential to further study of the NANB agent. Throughout these prospective studies, my main collaborator was Bob Purcell of the National Institute of Allergy and Infectious Diseases, who provided the basic research arm necessary to propel the investigations of NANB. From among the 50 NANB hepatitis cases identified at that time, I selected a patient (H) who had a particularly severe acute NANB hepatitis and from whom we had obtained apheresis units while his ALT levels were increasing. Purcell established a titration series of the H plasma and then did infectivity studies in the chimp. H had an infectivity titer of $10^{6.5}$ CID_{50}/ml (chimp infectious $doses_{50}$), allowing us to then undertake a series of manipulations of the virus and to test their effect on infectivity. Such studies by Feinstone[8], He[9] and others showed that the NANB agent was sensitive to chloroform and less than 30 nm in diameter, based on filtration. The agent of NANB hepatitis thus seemed to be small, lipid-enveloped, blood-transmissible and responsible for most residual cases of transfusion-associated hepatitis. Subsequently, Patrizia Farci from the University of Cagliari, in collaboration with Purcell and I, did a series of chimpanzee studies[10,11] in which she mixed chronic phase serum from patient H with the acute-phase infectious inoculum and studied the infectivity of the mixture in the chimp model. These studies made the important observation that neutralizing antibodies against hepatitis C virus (HCV) develop, but are very strain-specific and, in most cases, incapable of preventing the emergence of viral variants that lead to persistent infection.

As these virologic and immunologic studies were proceeding in the late 1970s and early 1980s, my main focus was to define the clinical consequences of NANB virus infection and to establish an assay that might be amenable to blood screening. The former proved easier than the latter. The entity NANB hepatitis initially met with considerable skepticism, and some believed it caused only an irrelevant transaminitis, because few patients had recognized clinical illness. However, as we monitored patients long-term and did liver biopsies in collaboration with Jay Hoofnagle and the NIH Liver Service, it became apparent that most NANB-infected patients had biochemical evidence of chronic hepatitis, and that 20% progressed to cirrhosis over the course of one to two decades. Later, after the discovery of HCV, we expanded these natural history studies in both asymptomatic donors[12] and transfusion recipients[13], the latter in collaboration with Leonard Seeff, and confirmed that 20–30% of HCV-infected individuals have severe histologic outcomes. However, equally important, these studies showed that about 20% of HCV-infected individuals undergo spontaneous recovery and that most have an indolent, perhaps non-progressive, course. In collaboration with Farci and Purcell, we have also shown the considerable viral diversity ('quasispecies') of HCV infection and that the extent of diversity in the acute phase of illness predicted whether chronic infection would ensue[14].

The histologic studies that documented progression to cirrhosis made it even more imperative to develop a blood-screening assay. In the decade from 1978 to 1988, we attempted every permutation of serologic approaches to assay development. Despite using 'highly pedigreed' infectious specimens, 'presumed convalescent' sera, eluted fractions, purified gamma globulins and the most-sensitive radio-immune assay approaches, we were unable to develop a specific serologic test for this elusive agent. In the absence of a specific assay, we looked for 'surrogate' markers that might identify NANB carriers. The most logical approach was measuring ALT. Although a retrospective analysis of our prospective collections showed that increases in ALT in the donor correlated with hepatitis transmission[15], we were unable to show the efficacy of this 'surrogate' assay in a subsequent prospective study. We then sought other measures of donor intervention and reasoned that donors who had been exposed to HBV might also be more likely to have been exposed to NANB; such donors were likely to have recovered from HBV infection and pass the donor screen, but might be persistent carriers of NANB. Thus, we used antibody against hepatitis B core antigen (HBc) as an index of past HBV infection, and showed in a retrospective analysis of our cohort that donors with antibody against HBc were four times more likely to transmit NANB hepatitis and that their exclusion might prevent 30% of such transmissions[16]. These data and those from a multicenter collaborative transfusion-transmitted viruses study[17] convinced the main blood organizations to implement testing for antibodies against HBc and for ALT in routine donor screening in 1987. It was difficult to measure the specific effect of these 'surrogate' assays because the threat of transfusion-associated AIDS had emerged and the 'surrogate' assays were introduced in concert with more-intensive questioning of donors regarding high-risk behavior and by a lessened use of allogeneic blood. Nonetheless, we could show that these combined measures served to decrease hepatitis incidence to 4.5% by 1989 (Fig. 2). Efforts to develop a specific NANB assay continued throughout the 1980s, although the main effort by Chiron was kept well concealed.

During this time, I had developed a panel of sera consisting of duplicate coded samples that had been proved to be infectious in the chimp or non-infectious in humans. By 1989, many different laboratories claimed to have developed a NANB assay and asked to test the panel. None was able to break the code and by 1989, the score was viruses, 20; investigators, zero. At that time, I received a call from George Kuo at Chiron, saying that they too felt they had a NANB assay. I sent George the remnants of the now-dwindling panel and within days received their results followed by several anxious calls asking if I had yet broken the code. When I did, I was excited to find that Chiron had detected all but two of the infectious sera and had properly found all the non-infectious sera to be negative. Further, the two samples that they missed were acute-phase sera, and subsequent samples from these same patients proved to be positive for what Chiron now called the hepatitis C virus. Michael Houghton will describe the events that preceded this discovery.

Using the newly developed assay for antibodies against HCV, we again delved into our repository and were able to rapidly show that 88% of NANB hepatitis cases seroconverted for antibody against HCV, that the development of antibody was in temporal relationship to the course of hepatitis and that infected patients could be linked to infected donors[18]. Thus, by 1990 it was clear that HCV was the principal agent of NANB hepatitis, and universal donor screening was initiated. We established a new prospective study to measure the effect of such testing and to define the extent of residual hepatitis unrelated to HBV or HCV. The first-generation assay for antibody against HCV resulted in a further 70% decrease in hepatitis incidence to a residual rate of 1.5%, and a more-sensitive second-generation assay, introduced in 1992, nearly eliminated HCV transmission (Fig. 2). Although mathematical modeling indicates that antibody-screened blood might still transmit HCV to 1:100,000 to 1:200,000 recipients, the observed decrease from 33% in 1970 to nearly zero in 1997 stands as a testament to the cumulative effectiveness of a series of donor screening interventions that were evidence-based. Viral nucleic acid testing of donors and improved viral inactivation technologies will soon bring transmission of hepatitis and human immunodeficiency virus from near-zero to absolute zero. I am now looking for another line of work.

Acknowledgments

Throughout almost the entire course of these clinical investigations, my right arm, and sometimes my left as well, has been my dedicated assistant, J. Melpolder. There is no way to adequately acknowledge the substantial contribution she has made in coordinating these studies that have involved thousands of patients. My gratitude is without bounds. I would also like to gratefully acknowledge the manifold contributions of my long-term associate James W. Shih, Ph.D. who so ably supervised the diverse laboratory aspects of these prospective studies.

The hepatitis C virus: A new paradigm for the identification and control of infectious disease

MICHAEL HOUGHTON

Identification of the hepatitis C virus

The problem of non-A, non-B (NANB) hepatitis emerged in 1975, after serological tests for hepatitis A virus (HAV) and hepatitis B virus (HBV) were developed. It then became evident that most hepatitis cases after transfusion were not due to either HAV or HBV, and that the risk of NANB hepatitis after blood transfusions was as high as 10% or even greater[5]. Later, it also became evident that NANB hepatitis occurred frequently in the form of sporadic, community-acquired infections. A frustrating period of 13 years followed, in which the methods successfully used to identify HAV and HBV all failed to result in the molecular identification of the etiological agent(s) of NANB hepatitis. No NANB-hepatitis-specific antigen, antibody or cell culture system was identified, and this lack of a molecular 'handle' limited progress in identifying the causative agent(s) of NANB hepatitis[19]. However, a chimpanzee model successfully developed by several groups[6,20,21] was exploited to show that one NANB hepatitis agent induced characteristic membranous tubules within the endoplasmic reticulum of infected chimpanzee hepatocytes[22]. Known as the tubule-forming agent, it was later shown to be filterable and to lose infectivity after treatment with organic solvents, consistent with its being a lipid-enveloped virus, possibly a toga-like or flavi-like virus[9,23]. Other data supported the existence of an HBV-like NANB hepatitis agent[19] as well as a chloroform-resistant (non-enveloped) NANB hepatitis virus[23] and possibly other NANB hepatitis agents[22].

Eventually, what turned out to be the main form of parenterally transmitted NANB hepatitis was identified using a 'blind' immunoscreening approach applied to recombinant λgt11 (ref. 24) cDNA libraries prepared from total RNA and DNA extracted from infectious chimpanzee plasma[25]. Serum from a patient diagnosed with NANB hepatitis was used as a presumed (but unproven) source of NANB-hepatitis-specific antibodies to identify just one viral cDNA clone from a complex cDNA library constituting one million other cDNAs. Formal proof of its etiological origin came from the demonstration that the clone was not derived from the host genome, that it bound a large RNA molecule of around 10,000 nucleotides found only in NANB-hepatitis-infectious materials, and that it was derived from a positive-stranded RNA encoding a protein that induced antibodies only in NANB hepatitis-infected individuals[25,39]. The RNA genome also encoded a large polyprotein of about 3,000 amino acids that had distant primary sequence identity with members of the Flaviviridae family[26]. HCV was therefore identified by direct molecular cloning of its genome in the relative absence of knowledge concerning the nature of the infectious agent and the immune response. This 'blind' method could be of value in the future in unearthing other unknown infectious agents involved in disease. The molecular identification of HCV was the culmination of a team effort[25] spanning 7 years, during which hundreds of millions of bacterial cDNA clones were screened for a putative NANB hepatitis origin using many different approaches. The successful approach involved Qui-Lim Choo in my Laboratory at Chiron and the laboratories of George Kuo (Chiron) and Daniel Bradley (CDC). I accept the award on behalf of these collaborators (Fig. 3).

Figure 3 The HCV team (from left to right; M. Houghton, Q-L Choo, G. Kuo and D. Bradley.

Properties of HCV

Now known to contain a highly variable RNA genome, HCVs constitute a large genus (the hepacivirus genus) within the Flaviviridae family. Six basic genotypes have been distinguished so far, with more than 100 phylogenetically-distinct subtypes[27]. At any one time, the viral genome exists as a complex quasispecies. The RNA genome contains a conserved 5′-terminal internal ribosome entry site that is responsible for initiating translation of the large polyprotein. The latter is cleaved co- and post-translationally into at least three structural or virion proteins and seven presumed non-structural proteins involved in replication of the virus[28-30] (Fig. 4). The 3′ terminus of the RNA genome is composed of a variable region, a polypyrimidine tract and a highly conserved stem–loop secondary structure[29]. Hypervariable regions exist within the large gpE2 glycoprotein domain that may be under immune selection[31]. The virus cannot be grown efficiently in cell culture or purified from infected liver or blood, and thus still has not been characterized morphologically or biochemically.

Serodiagnosis

The molecular cloning of the HCV genome led to the availability of many recombinant HCV diagnostic antigens. George Kuo purified these and developed numerous experimental EIA tests detecting HCV antibodies. This intensive work allowed the selection of optimal, immunodominant epitopes for inclusion in an evolving series of sensitive and specific blood screening and diagnostic tests for HCV infection[32]. Used to screen blood donors beginning in 1990, these assays have led to the near-disappearance of transfusion-associated hepatitis C. At least 40,000 infections have been prevented each year in the US alone since the implementation of these tests[33]. Such tests have also been of great use in diagnosing hepatitis patients and in their clinical management, and in attributing liver and extra-hepatic diseases to HCV infection. Chronic hepatitis, liver cirrhosis, hepatocellular carcinoma, cryoglobulinemia and porphyria cutanea tarda are all well-established potential clinical sequelae of chronic persistent HCV infection. Other diseases, such as oral lichen planus, Sjögren's-like syndrome and non-Hodgkin lymphoma, have also been linked with HCV infection[28]. Although most HCV-infected individuals have few clinical symptoms and will not progress to a severe disease state, a subset can undergo progressive liver disease in which, often over decades, chronic hepatitis develops into liver cirrhosis and then into hepatocellular carcinoma. Extra-hepatic manifestations can also occur.

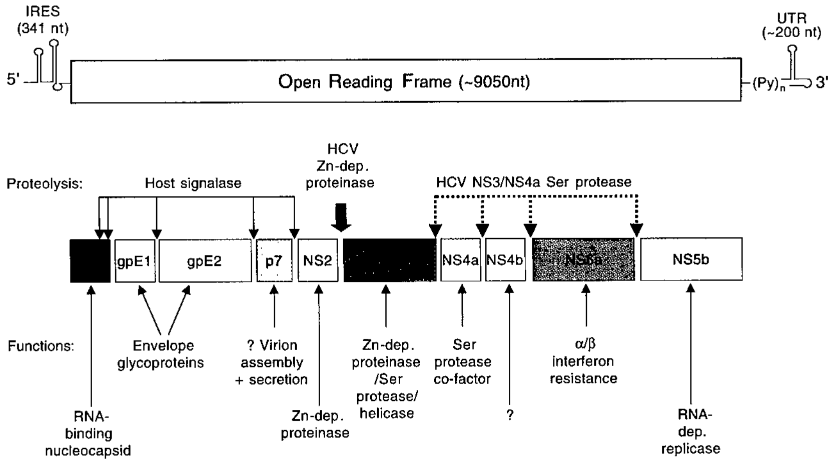


Figure 4 The organization of the hepatitis C positive-stranded RNA genome.

Recently, various nucleic acid testing assays have been developed, based on the highly conserved 5′ internal ribosome entry site and nucleocapsid gene sequences, which can diagnose new infections before seroconversion to having antibodies against HCV occurs. Implementation of these tests for screening of blood donors will reduce the risk of HCV after transfusion still further (down to approximately one in 300,000 in the US, for example). A test detecting circulating nucleocapsid antigen is also of value in diagnosing infection before seroconversion.

Education, new therapies and vaccines

Future challenges exist for both developed and developing countries. In the latter, global implementation of blood donor screening for HCV has been recommended recently by the World Health Organization. In many countries in which HCV is endemic, the risks of HCV infection after transfusion are still exceedingly high. Also, the World Health Organization has emphasized education to lower the risks of HCV transmission in both developing and developed countries. The historical use of non-sterile injection devices has been mainly responsible for the huge burden of HCV disease present in many developing countries, as well as cultural practices (such as circumcision) involving the use of non-sterile medical equipment. In developed countries, intravenous drug use involving sharing of needles/syringes is still the main risk factor[34]. Any procedure involving blood transfer (such as tattooing using shared instruments) is not recommended. The Centers for Disease Control also recommends that HCV-infected individuals not share toothbrushes, razors and so on.

The current therapy for HCV consists of a combination of interferon and ribovirin[35]. However, both drugs can produce substantial toxicity, and only a minority of patients responds. In particular, long-term response rates with the most common genotype, type 1, occur in only approximately 30% of patients. Although the imminent introduction of a more stable form of pegylated interferon will improve response rates somewhat, it is apparent that more-effective and less-toxic drugs are required. The HCV genome encodes two proteases involved in processing of the viral polyprotein: a helicase involved in unwinding the RNA strands during replication and translation, and a replicase that copies the positive RNA strand (Fig. 4). The fine structures of all these enzymes have now been resolved by X-ray diffraction methods and are now the subjects of rational drug design[29,30].

Other therapeutic developments involve ribozyme and antisense strategies involving the conserved 5′ internal ribosome entry site and nucleocapsid gene sequences, and nucleoside analog inhibitors of the replicase. A putative receptor for HCV, the CD81 tetraspanin molecule, is also the subject of potential antiviral development[36]. Therefore, we can be optimistic that new, specific drugs against HCV will emerge within the next 5–10 years.

Between 12 and 50% of acute HCV infections spontaneously resolve without progressing to the chronically infected state that is associated with the pathogenic sequelae of infection. Such resolution of acute infection is associated with the induction of broad helper and cytotoxic T-lymphocyte responses to the virus[37]. Thus, appropriate vaccination to prime such immune responses may lower the high chronicity rate associated with HCV infection. Vaccination of chimpanzees with recombinant envelope glycoproteins gpE1 plus gpE2 successfully prevented the development of chronic infection in most animals after challenge with homologous or heterologous subtype 1a virus. In contrast, most control unvaccinated animals develop chronic, persistent infections[37]. These data are encouraging for the development of human vaccines. Human immunoglobulin preparations (containing antibodies against HCV) have been reported to be effective at reducing the rate of development of chronic infection in the transfusion setting, between sexual partners and in liver transplantation[37]. Response to interferon has also been linked with the endogenous level of intrahepatic HCV-specific cytotoxic T-lymphocyte activity before treatment[38]. If it is confirmed, appropriate vaccination may also be important therapy if such virus-specific cytotoxic T-lymphocyte activity can be boosted during drug therapy.

Finally, although HCV is an RNA virus that does not produce DNA replication intermediates that can integrate into the host genome, it still manages to persist in most cases (usually for life in untreated individuals). Although it may involve the emergence of viral escape mutants to both antibody and T-cell responses, it is very likely that additional mechanisms are in operation to result in such high rates of chronicity. Elucidating these mechanisms represents the most intriguing challenge of future HCV research and, in the process, is likely to open up new strategies for the control of this challenging virus.

Acknowledgements

I gratefully acknowledge the input of the late Lacy Overby as well as Amy Weiner in this work.

1. Blumberg, B.S., Alter, H.J.& Visnich, S. A "new" antigen in leukemia sera. *JAMA* **191**, 541–546 (1965).
2. Alter, H.J.& Blumberg, B.S. Further studies on a "new" human isoprecipitin system (Australia antigen). *Blood* **27**, 297–309 (1966).
3. Alter, H.J. *et al.* Posttransfusion hepatitis after exclusion of the commercial and hepatitis B antigen positive donor. *Ann. Int. Med.* **77**, 691–699 (1972).
4. Feinstone, S.M., Kapikian, A.Z. & Purcell, R.H. Hepatitis A detection by immune electron microscopy of a virus-like antigen associated with acute illness. *Science* **182**, 1026–1028 (1973).
5. Feinstone, S.M., Kapikian, A.Z., Purcell, R.H., Alter, H.J. & Holland, P.V. Transfusion-associated hepatitis not due to viral hepatitis type A or B. *N. Engl. J. Med.* **292**, 767–770 (1975).
6. Alter, H.J., Purcell, R.H., Holland, P.V. & Popper, H. Transmissible agent in "non-A, non-B" hepatitis. *Lancet* **1**, 459–463 (1978).
7. Alter, H.J. *et al.* Transmission of HTLV-III Infection from human plasma to chimpanzees: an animal model for the acquired immunodeficiency syndrome. *Science* **226**, 549–552 (1984).
8. Feinstone, J.M. *et al.* Inactivation of hepatitis b virus and non-a, non-b virus by chloroform. *Infect. Immun.* **4**, 816–821 (1983).
9. He, L-F. *et al.* Determining the size of non-A, non-B hepatitis virus by filtration. *J. Inf. Dis.* **156**, 636–640 (1987).
10. Farci, P. *et al.* Lack of protective immunity against reinfection with hepatitis C virus. *Science* **258**, 135–140 (1992).
11. Farci, P. *et al.* Prevention of HCV infection in chimpanzees following antibody-mediated in vitro neutralization. *Proc. Natl. Acad. Sci. USA* **91**, 7792 (1994).
12. Conry-Cantilena, C. *et al.* Routes of infection, viremia, and liver disease in blood donors found to have hepatitis C virus infection. *N. Engl.J. Med.* **334**, 1691–1696 (1996).
13. Seeff, L.B. *et al.* Long-term mortality after transfusion-associated non-A, non-B hepatitis. *N .Engl. J. Med.* **327**, 1906–1911 (1992).
14. Farci, P. *et al.* The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* **288**, 339–344 (2000).
15. Alter, H.J., Purcell, R.H., Holland, P.V., Alling, D.W. & Koziol, D.E. The relationship of donor transaminase (ALT) to recipient hepatitis: impact on blood transfusion services. *JAMA* **246**, 630–634 (1981).
16. Koziol, D.E. *et al.* Antibody to hepatitis B core antigen as a paradoxical marker for non-A, non-B hepatitis agents in donated blood. *Ann. Intern. Med.* **104**, 488–495 (1986).
17. Stevens, C.E. *et al.* Hepatitis B virus antibody in blood donors and the occurrence of non-A, non-B hepatitis in transfusion recipients: an analysis of the transfusion-transmitted viruses study. *Ann. Intern. Med.* **104**, 488–495 (1984).
18. Alter, H.J. *et al.* Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. *N. Engl. J. Med.* **321**, 1494–1500 (1989).
19. Shih, J.W.-K, Esteban, J.I. & Alter, H.J. Non-A, non-B hepatitis: advances and unfulfilled expectations of the first decade. *Prog. Liver Dis.* **8**, 433–452 (1986).
20. Tabor, E. *et al.* Transmission of non-A, non-B hepatitis from man to chimpanzee. *Lancet* 1, 463–466 (1978).
21. Hollinger, F.B., Gitnick, G.L. & Aach, R.D. Non-A, non-B hepatitis transmission in chimpanzees:a project of the transfusion-transmitted viruses study group. *Intervirology* **10**, 60–68 (1978).
22. Shimizu, Y.K. *et al.* Non-A, non-B hepatitis: ultrastructural evidence for two agents in experimentally infected chimpanzees. *Science* **205**, 197–200 (1979).
23. Bradley, D.W. The agents of non-A, non-B hepatitis. *J. Virol. Meth.* **10**, 307–319 (1985).
24. Young, R.A. & Davis, R.W. Efficient isolation of genes by using antibody probes. *Proc. Natl. Acad.Sci. USA* **80**, 1194–1198 (1983).
25. Choo, Q.-L. *et al.* Isolation of a cDNA clone derived from a blood-borne non-A,non-B viral hepatitis genome. *Science* **244**, 359–362 (1989).
26. Choo, Q.L. *et al.* Genetic organisation and diversity of the hepatitis C virus. *Proc. Natl. Acad. Sci. USA* **88**, 2451–2455 (1991).
27. Simmonds, P. Viral heterogeneity of the hepatitis C virus. *J. Hepatol.* **31**, 54–60 (1999).
28. Houghton, M. in *Fields Virology* (eds. Fields, B.N. *et al.*) 1035–1058 (Raven, Philadelphia, 1996).
29. Reed, K. & Rice, C. in *The Hepatitis C Viruses* (eds. Hagedorn, C. & Rice, C.) 55-84 (Springer, Berlin, 2000).
30. DeFrancesco, R. *et al.* Biochemical and immunologic properties of the nonstructural proteins of the hepatitis C virus: implications for development of antiviral agents and vaccines. *Semin. Liver Dis.* **20**, 69–83 (2000).
31. Weiner, A.J. *et al.* Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. *Proc. Natl. Acad. Sci. USA* **89**, 3468–3472 (1992).
32. Kuo, G. *et al.* An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science* **244**, 362–364 (1989).
33. Alter, H.J. Discovery of non-A, non-B hepatitis and identification of its etiology. *Am. J. Med.* **107**, 16S–20S (1999).
34. Wasley, A. & Alter, M. Epidemiology of hepatitis C: geographical differences and temporal trends. *Semin. Liver Dis.* **20**, 1–16 (2000).
35. Davis, G.L. Current therapy for chronic hepatitis C. *Gastroenterology* **118**, S104–S114 (2000).
36. Pileri, P. *et al.* Binding of hepatitis C virus to CD81. *Science* **282**, 938–941 (1998).
37. Houghton, M. in *The Hepatitis C Viruses* (eds. Hagedorn, C. & Rice, C.) 327–339 (Springer, Berlin, 2000).
38. Nelson, D.R., Marousis, C.G., Ohno, T., Davis, G.L. & Lau, J.Y. Intrahepatic hepatitis C virus-specific cytotoxic T lymphocyte activity and response to interferon alpha therapy in chronic hepatitis C. *Hepatology* **28**, 225–230 (1998).
39. Ezzell, C. Candidate cause identified of non-A, non-B hepatitis. *Nature* **333**, 195 (1998).

Harvey J. Alter
Chief, Immunology Section, Department of Transfusion Medicine
Warren Grant Magnuson Clinical Center
NIH Building 10, Room 1C711, 9000 Rockville Pike
Bethesdsa, Maryland 20892

Michael Houghton
Chiron Corporation
4560 Horton Street
Emeryville, CA 94608-2916

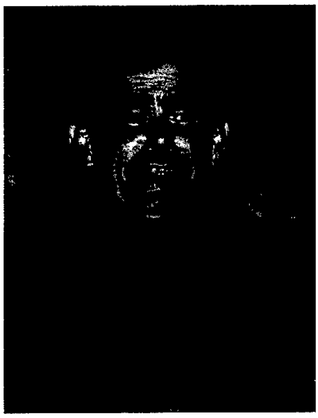
Special Achievement in Medical Science Award

From cell physiology to cell physiology

SYDNEY BRENNER

I was born in Germiston, a small town outside Johannesburg, South Africa, on 13 January 1927. Most of my early education was obtained from the local Carnegie Public Library, where I first learned how to acquire knowledge from books—something that, unfortunately, has almost completely disappeared from the modern world. A bursary from the local town council enabled me to enter medical studies at the University of Witwatersrand in 1942. I deviated into a medical science course after two years, when it was discovered that I would be too young to qualify as a doctor at the end of the six-year course. I did two additional years of medical sciences, and my MSc thesis was on the chromosomes of *Elephantulus*, an insectivore. I taught myself this field by reading Darlington's *Recent Advances in Cytology*, and I became a skilled microscopist and histologist both from my work and from working part-time as a technician in the Anatomy Department.

When I began my research, I already had several years of post-graduate experience in 'garage' chemistry and I had added some biochemistry to my interests. I decided that cells were the important biological entities to study, particularly living cells, and the subject I decided to take up was one I called cell physiology. I began to build various pieces of equipment that I thought I would need. One was a Warburg respirometer, but my glassblowing was not very inspired. The other, more successful, was a small Beams-and-King air turbine ultracentrifuge. I used this to produce the first paper[1] I was really proud of and still am. In it, I showed that the material called 'chromidial substance', which stained red with methyl green pyronin and contained RNA, corresponded to the fraction that Claude had separated from ground-up cells and called microsomes. These were characterized by having certain sedimentation properties in extracts. I showed, by centrifuging small pieces of rat liver at 300,000*g* in my small turbine, that the material identified cytologically had the same properties in each cell as the bulk material in the extract.


Mark Ewing Photography, St. Ives, Huntingdon, Cambs PE17 4JH

Sydney Brenner.

I returned to complete my medical studies in 1947, but I was not an exemplary medical student. Once, when a spherical thoracic surgeon uttered the remark that surgery was an exact science like physics and chemistry, I burst into uncontrollable hysterical laughter and was ordered to leave. I never returned to that part of the course, and so that part of the body remained a mystery. Of course, as a skilled human anatomist, I knew and still know what is in the thorax, but I never learned how to open it.

At the same time that I was studying medicine as well as researching and teaching, I continued my fanatical collection of knowledge, nearly all self-taught. When I read that solutions of DNA were thixotropic, I studied rheology, but was disappointed because the subject seemed to have more to do with the food industry and how you make porridge than with genes and cells. I also tried to prepare myself for the future by learning mathematics, but it was not clear what would be needed and so I learned a little of everything. I continued my earlier studies in chemistry by synthesizing dyes, and published three papers on supravital staining[2–4].

All of this prepared me for a career in molecular biology, but the subject had still not been invented when I came to Oxford in 1952 to study bacteriophage in the physical chemistry laboratory. I worked with Sir Cyril Hinshelwood, who had written a book called *The Physical Chemistry of the Bacterial Cell*, which I thought was close enough to the yet-unborn subject. I had my own ideas about how DNA might be related to proteins (these were wrong), and I thought we could study the structure of DNA by optical methods with acridine dyes, a topic later revisited in a different form[5,6]. Jack Dunitz and Leslie Orgel, whom I met at Oxford, eventually led me to the Cavendish Laboratory in Cambridge. There, in April 1953, I met Francis Crick and James Watson and the DNA model. I knew then and there exactly what I was going to do for the rest of my life.

I have written elsewhere about the exhilaration of belonging, in those early days of molecular biology, to a radical evangelical sect[7]. It was composed mostly of physicists, and I was the only failed thoracic surgeon among them. Like most of my colleagues I read Schroedinger's *What is Life*, but in 1946, I didn't understand it. After reading von Neumann's work on self-reproducing machines in 1952, I knew where Schroedinger had gone wrong. Schroedinger thought that the genetic material contained both the programme for development as well as the means for its execution, whereas von Neumann showed that it contains only a description of the means for executing the programme (see ref. 8).

For the last 50 years or so, all my scientific work has been directed toward determining how the genetic components in biological systems are implemented to generate organisms with complex structures and functions. This remains the central problem of biology, and is the same whether one studies bacteriophage or bacteria, or nematodes, fish or human beings. It is the same whether one does classical experimental genetics or one sequences genomes. Today, when we have organized large groups to sequence the human genome, and high throughput is on everyone's lips, and when we are puzzled by how we are going to integrate all of this information, I return to the thought I had quite early that genes are not the units of development or function. Cells are the units, and our job is to discover how the products of the genes work in cells to govern their activities.

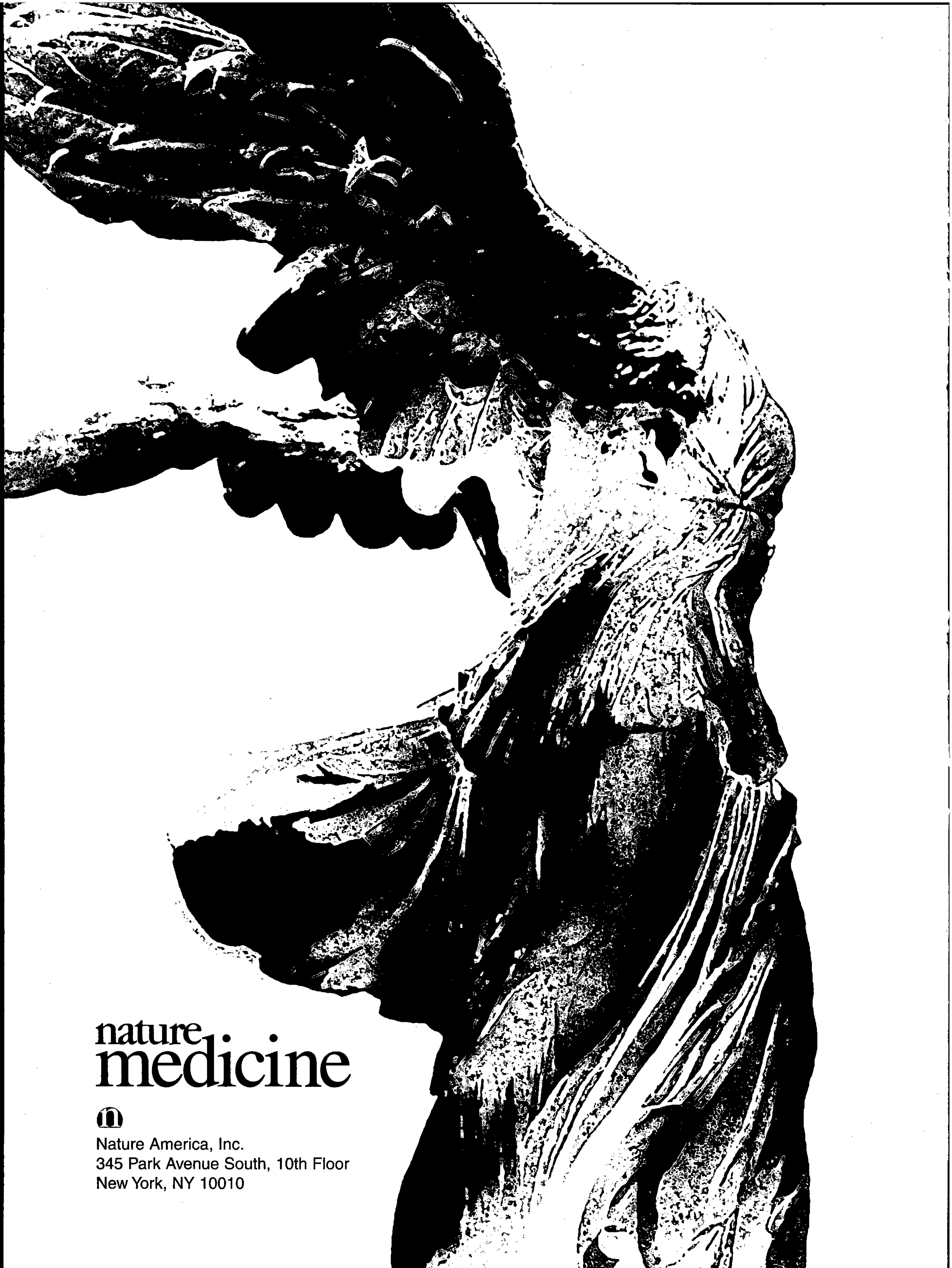
I wrote the following as an introduction to a Ciba

Symposium in 1978 (ref. 9): "The great ease with which molecular information can be collected on the genomes of higher organisms will tempt many. We can inevitably expect vast compendia of sequences but, without functional reference, these compendia will be uninterpretable, like an undeciphered ancient language. Many people and many computers will play games with these sequences, but we will have to find out by experiment what the sequences do and how the products they make participate in the physiology and development of the organism. Thus, although the analysis of the genotype has been taken care of, we still need better ways of analyzing phenotypes. Many of us are ultimately interested in the causal analysis of development and the reduction of the complex phenotypes of higher organisms to the level of gene products. This is still the major problem of biology. We must understand what cells can do because all of what we are is generated by cells growing, moving and differentiating."

This is still valid today.

1. Brenner, S. The identity of the microsomal lipoprotein-ribonucleic acid complexes with cytologically observable chromidial substance (cytoplasmic ribonucleoprotein) in the hepatic cell. *S. Afr. J. Med. Sci.* **12**, 53–60 (1947).
2. Brenner, S. The demonstration by supravital dyes of oxidation-reduction systems on the mitochondria of the rat lymphocyte. *S. Afr. J. Med. Sci.* **14**, 13–19 (1949).
3. Brenner, S. Supravital staining of mitochondria with amethyst violet. *Stain Technol.* **25**, 163–164 (1950).
4. Brenner, S. Supravital staining of mitochondria with phenosafranin dyes. *Biochim. Biophys. Acta* **11**, 480–486 (1953).
5. Brenner, S., Benzer, S.& Barnett, L. Distribution of proflavin-induced mutations in the genetic fine structure. *Nature* **182**, 983–985 (1958).
6. Crick, F.H.C., Barnett, L., Brenner, S. & Watts-Tobin, R.J. General nature of the genetic code for proteins. *Nature* **192**, 1227–1232 (1961).
7. Brenner, S. New directions in molecular biology. *Nature* **248**, 785–787 (1974).
8. von Neumann, J. "The general and logical theory of automata" in *Cerebral Mechanisms in Behavior* (ed. Jefress, L.) 1–31 (Hafner, New York 1951).
9. Brenner, S. Human genetics: possibilities and relatives, Ciba Foundation Symposium. *Excerpta Medica* **66**, 1–3 (1973).

Sydney Brenner
The Molecular Sciences Institute
2168 Shattuck Avenue, 2nd Floor
Berkeley, CA 94704



nature
medicine

Nature America, Inc.
345 Park Avenue South, 10th Floor
New York, NY 10010

osmium tetroxide in 0.05 M cacodylate buffer pH 7.0 at 4 °C for 2 h, rinsed in 0.025 M sodium phosphate (pH 7.0), dehydrated in a graded ethanol series at 4 °C and critical-point-dried in a Polaron E3000 Critical Point Drier using CO_2. Seedlings were mounted on stubs, and organs obscuring the apical meristem were removed. Specimens were coated with 20 nm of gold–palladium in a Polaron E5000 Sputter Coater, and examined on a Philips XL30FEG scanning electron microscope at 5 kV.

Statistical analysis

Statistical tests were performed using GraphPad Instat Version 3.0 (GraphPad Software).

Received 12 November 1999; accepted 23 March 2000.

1. Lyndon, R. F. *The Shoot Apical Meristem* (Cambridge Univ. Press, Cambridge, 1998).
2. Clark, S. E. & Schiefelbein, J. W. Expanding insights into the role of cell proliferation in plant development. *Trends Cell Biol.* **7,** 454–458 (1997).
3. Kinsman, E. A. *et al.* Elevated CO_2 stimulates cells to divide in grass meristems: a differential effect in two natural populations of *Dactylis glomerata. Plant Cell Envir.* **20,** 1309–1316 (1997).
4. Francis, D. in *Inherent Variation in Plant Growth* (eds Lambers, H. & Van Vuuren, M. M. I.) 5–20 (Backhuys, Leiden, 1998).
5. Francis, D. & Barlow, P. W. Temperature and the cell cycle. *Symp. Soc. Exp. Biol.* **42,** 181–201 (1988).
6. Hagemann, W. The relationship of anatomy to morphology in plants: A new theoretical perspective. *Int. J. Plant Sci.* **153,** 538–548 (1992).
7. Barlow, P. W. in *Shape and Form in Plants and Fungi* (eds Ingram, D. S. & Hudson, A.) 169–193 (Academic, London, 1995)
8. Soni, R., Carmichael, J. P., Shah, Z. H. & Murray, J. A. H. A family of cyclin D homologs from plants differentially controlled by growth regulators and containing the conserved retinoblastoma protein interaction motif. *Plant Cell* **7,** 85–103 (1995).
9. Bayliss, M. W. in *The Cell Division Cycle in Plants* (eds Bryant, J. A. & Francis, D.) 157–177 (Cambridge Univ. Press, Cambridge, 1985).
10. Pines, J. Cyclins and cyclin-dependent kinases: Theme and variations. *Adv. Cancer Res.* **66,** 181–212 (1995).
11. Murray, J. A. H. *et al.* in *Plant Cell Division* (eds Francis, D., Dudits, D. & Inzé, D.) 99–127 (Portland, London, 1998).
12. Riou-Khamlichi, C., Huntley, R., Jacqmard, A. & Murray, J. A. H. Cytokinin activation of *Arabidopsis* cell division through a D-type cyclin. *Science* **283,** 1541–1544 (1999).
13. Setiady, Y. Y., Sekine, M., Hariguchi, N., Kouchi, H. & Shinmyo, A. Molecular cloning and characterization of a cDNA clone that encodes a *cdc2* homolog from *Nicotiana tabacum. Plant Cell Physiol.* **37,** 369–376 (1996).
14. Mironov, V., De Veylder, L., Van Montagu, M. & Inzé, D. Cyclin-dependent kinases and cell division in plants—the nexus. *Plant Cell* **11,** 509–521 (1999).
15. Lyndon, R. F. Rates of cell division in the shoot apical meristem of *Pisum. Ann. Bot.* **34,** 1–17 (1970).
16. Galli, M. G. & Sala, F. Aphidicolin as synchronizing agent in root tip meristems of *Haplopappus gracilis. Plant Cell Rep.* **2,** 156–159 (1983).
17. Francis, D., Davis, M. S., Braybrook, C., James, N. C. & Herbert, R. J. An effect of zinc on M-phase and G1 of the plant cell cycle in the synchronous TBY-2 tobacco cell suspension. *J. Exp. Bot.* **46,** 1887–1894 (1995).
18. Polymenis, M. & Schmidt, E. V. Coordination of cell growth with cell division. *Curr. Opin. Genet. Dev.* **9,** 76–80 (1999).
19. Gonthier, R., Jacqmard, A. & Bernier, G. Occurrence of two cell subpopulations with different cell-cycle durations in the central and peripheral zones of the vegetative shoot apex of *Sinapis alba L. Planta* **165,** 288–291 (1985).
20. Clowes, F. A. L. in: *Cell Division in Higher Plants* (ed. Yeoman, M. M.) 253–284 (Academic, London, 1976).
21. Gonthier, R., Jacqmard, A. & Bernier, G. Changes in cell-cycle duration and growth fraction in the shoot meristem of *Sinapis* during floral transition. *Planta* **170,** 55–59 (1987).
22. Evans, H. J., Neary, G. J. & Tonkinson, S. M. The use of colchicine as an indicator of mitotic rate in broad bean root meristems. *J. Genet.* **55,** 487–502 (1957).
23. Neufeld, T. P., de la Cruz, A. F. A., Johnston, L. A. & Edgar, B. A. Coordination of growth and cell division in the *Drosophila* wing. *Cell* **93,** 1183–1193 (1998).
24. Nash, R., Tokiwa, G., Anand, S., Erickson, K. & Futcher, A. B. The *WHI1'* gene of *Saccharomyces cerevisiae* tethers cell division to cell size and is a cyclin homolog. *EMBO J.* **7,** 4335–4346 (1988).
25. Quelle, D. E. *et al.* Overexpression of mouse D-type cyclins accelerates G_1 phase in rodent fibroblasts. *Genes Dev.* **7,** 1559–1571 (1993).
26. Resnitzky, D., Gossen, M., Bujard, H. & Reed, S. I. Acceleration of the G_1/S phase transition by expression of cyclins D1 and E with an inducible system. *Mol. Cell. Biol.* **14,** 1669–1679 (1994).
27. Doerner, P., Jorgensen, J-E., You, R., Steppuhn, J. & Lamb, C. Control of root growth and development by cyclin expression. *Nature* **380,** 520–523 (1996).
28. Hemerly, A. *et al.* Dominant negative mutants of the Cdc2 kinase uncouple cell division from iterative plant development. *EMBO J.* **14,** 3925–3936 (1995).
29. Gleave, A. P. A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Mol. Biol.* **20,** 1203–1207 (1992).
30. Riou-Khamlichi, C., Menges, M., Healy, J. M. S. & Murray, J. A. H. Sugar control of the plant cell cycle: Differential regulation of *Arabidopsis* D-type cyclin gene expression. *Mol. Cell. Biol.* (in the press).

Acknowledgements

We thank M. Sekine for permission to cite unpublished data; I. Furner for making possible the glasshouse experiments; R. Day and A. Mills for expert plant care and measurements; A. Inskip for technical support, T. Burgess and the Multi-Imaging Centre for help with scanning electron micrographs; and M. Cornelissen and M. De Block for invaluable discussions. This work was supported in part by the UK Biotechnology and Biological Sciences Research Council.

Correspondence and requests for materials should be addressed to J.A.H.M. (e-mail: j.murray@biotech.cam.ac.uk).

Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway

Glenn C. Turner*, Fangyong Du* & Alexander Varshavsky

Division of Biology, California Institute of Technology, Pasadena, California 91125, USA

** These authors contributed equally to this work*

Protein degradation by the ubiquitin system controls the intracellular concentrations of many regulatory proteins. A protein substrate of the ubiquitin system is conjugated to ubiquitin through the action of three enzymes, E1, E2 and E3, with the degradation signal (degron) of the substrate recognized by E3 (refs 1–3). The resulting multi-ubiquitylated substrate is degraded by the 26S proteasome[4]. Here we describe the physiological regulation of a ubiquitin-dependent pathway through allosteric modulation of its E3 activity by small compounds. Ubr1, the E3 enzyme of the N-end rule pathway (a ubiquitin-dependent proteolytic system) in *Saccharomyces cerevisiae* mediates the degradation of Cup9, a transcriptional repressor of the peptide transporter Ptr2 (ref. 5). Ubr1 also targets proteins that have destabilizing amino-terminal residues[6]. We show that the degradation of Cup9 is allosterically activated by dipeptides with destabilizing N-terminal residues. In the resulting positive feedback circuit, imported dipeptides bind to Ubr1 and accelerate the Ubr1-dependent degradation of Cup9, thereby de-repressing the expression of Ptr2 and increasing the cell's capacity to import peptides. These findings identify the physiological rationale for the targeting of Cup9 by Ubr1, and indicate that small compounds may regulate other ubiquitin-dependent pathways.

The rate of degradation of specific proteins is often regulated by modulating the exposure or the structure of their degrons. For example, the degrons of the cyclin-dependent kinase inhibitors Sic1 and p27 are activated by phosphorylation, which is timed to bring about their destruction at key transition points in the cell cycle[7]. In other cases, phosphorylation regulates the activity of an E3 itself. For example, the anaphase-promoting complex, a multisubunit E3, is activated only at mitosis[8]. One ubiquitin-dependent proteolytic system, called the N-end rule pathway, targets proteins carrying a degradation signal called the N-degron[9], which comprises a destabilizing N-terminal residue and a lysine residue[10,11]. In *S. cerevisiae*, there are two classes of destabilizing residues, basic or 'type 1' (Arg, Lys and His), and bulky hydrophobic or 'type 2' (Phe, Leu, Tyr, Trp and Ile). Ubr1, a RING-H2 finger-containing E3 of relative molecular mass 225,000 (M_r 225K), directly recognizes these N-terminal residues[6,12]. A complex of Ubr1 and the E2 enzyme Rad6 (Ubc2) synthesizes a multi-ubiquitin chain linked to a lysine residue of the substrate[13]. Dipeptides with a destabilizing N-terminal residue of either basic or hydrophobic type act as competitive inhibitors of the degradation of N-end rule substrates carrying the same type of destabilizing residue[14–16]. Thus, Ubr1 contains two distinct N-terminal residue-binding sites that are each capable of binding either a dipeptide or a protein, but not both at the same time.

The N-end rule pathway was discovered through the use of engineered reporter proteins[9]. The first physiological function of Ubr1 in *S. cerevisiae* has recently been shown to be the regulation of peptide uptake[17], through control of degradation of the M_r 35K homeodomain protein Cup9, a transcriptional repressor of the di- and tripeptide transporter Ptr2 (ref. 5). Ubr1 targets Cup9 through a degron located in the carboxy-terminal half of Cup9 (F. Navarro-Garcia, G.C.T. and A.V., unpublished data). Despite this unexpected mode of recognition, we asked whether dipeptides bearing destabilizing N-terminal residues could affect the Ubr1-mediated degradation of Cup9, as dipeptides are able to inhibit degradation of canonical N-end rule substrates[16].

We addressed this question using Cup9$_{\mathrm{NSF}}$, a C-terminally Flag-tagged variant of the Cup9 repressor bearing an Asn 265→Ser substitution. Cup9$_{\mathrm{NSF}}$ was degraded indistinguishably from wild-type Cup9 but was predicted to have a much lower affinity for DNA[18,19], so it would not influence the expression of Ptr2 and the uptake of dipeptides. Cup9$_{\mathrm{NSF}}$ was expressed as part of a fusion of the form Flag–DHFR–ubiquitin–Cup9$_{\mathrm{NSF}}$ (where Flag–DHFR is N-terminally Flag-tagged mouse dihydrofolate reductase; see Methods; Fig. 1). Ubiquitin-specific proteases co-translationally cleave this UPR (ubiquitin–protein–reference) fusion at the ubiquitin–Cup9 junction, yielding the test protein Cup9$_{\mathrm{NSF}}$ and the long-lived Flag–DHFR–ubiquitin reference protein, which serves as an internal control for variations in expression levels and immunoprecipitation efficiency[20,21].

Cells expressing Cup9$_{\mathrm{NSF}}$ were grown in minimal medium containing allantoin as the nitrogen source to avoid the known effects of nitrogen catabolite repression on *PTR2* expression[22]. Leu–Ala and Arg–Ala, dipeptides bearing either type of destabilizing N-terminal residue (Leu, bulky hydrophobic; Arg, basic) were added to a final concentration of 10 mM (see Methods). This dipeptide concentration results in maximal inhibition of degradation of N-end rule substrates[16]. Notably, the addition of either Leu–Ala or Arg–Ala exerted an opposite effect on Cup9$_{\mathrm{NSF}}$, strongly accelerating its degradation in wild-type (*UBR1*) cells. The half-life of Cup9$_{\mathrm{NSF}}$ decreased from ~5 min in the absence of dipeptides (Fig. 1c) to less than 1 min in their presence (Fig. 1b). This stimulatory effect was not observed in a *ubr1Δ* strain, indicating that the augmented degradation of Cup9$_{\mathrm{NSF}}$ was dependent on Ubr1. The enhancement of degradation required dipeptides with destabilizing N-terminal residues; dipeptides of the same composition but bearing a stabilizing residue (Ala–Leu and Ala–Arg) did not affect the degradation of Cup9$_{\mathrm{NSF}}$ ($t_{1/2}$ ~5 min) (Fig. 1b, and data not shown). Similar results were obtained with cells expressing Cup9$_{\mathrm{NSF}}$ that was not a part of a UPR fusion (data not shown).

To determine the concentration dependence of the stimulation, we measured the degradation of Cup9 at a range of concentrations of Trp–Ala. The enhancement of Cup9$_{\mathrm{NSF}}$ degradation was detectable at 1 μM Trp–Ala, the lowest concentration tested ($t_{1/2}$ ~1 min) (Fig. 1c). In contrast, the degradation of Cup9$_{\mathrm{NSF}}$ was not altered by either Ala–Trp or the constituent amino acids Trp and Ala (Fig. 1c). Similar results were obtained using Leu–Ala and Arg–Ala (data not shown). These results indicate that the relevant signalling molecule in this process is a dipeptide bearing a destabilizing N-terminal residue. The range of dipeptide concentrations that significantly stimulated Cup9 degradation was similar to physiologically active levels of many other nutrients.

Cup9 represses transcription of the transporter-encoding *PTR2*

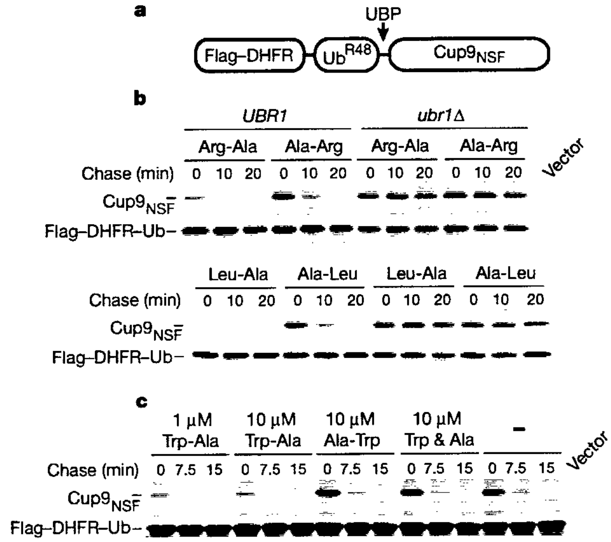

a
UBP
Flag–DHFR
Ub$^{\mathrm{R48}}$
Cup9$_{\mathrm{NSF}}$
b
UBR1
ubr1Δ
Arg-Ala
Ala-Arg
Arg-Ala
Ala-Arg
Vector
Chase (min) 0 10 20 0 10 20 0 10 20 0 10 20
Cup9$_{\mathrm{NSF}}$
Flag–DHFR–Ub
Leu-Ala
Ala-Leu
Leu-Ala
Ala-Leu
Chase (min) 0 10 20 0 10 20 0 10 20 0 10 20
Cup9$_{\mathrm{NSF}}$
Flag–DHFR–Ub
c
1 μM Trp-Ala
10 μM Trp-Ala
10 μM Ala-Trp
10 μM Trp & Ala
–
Vector
Chase (min) 0 7.5 15 0 7.5 15 0 7.5 15 0 7.5 15 0 7.5 15
Cup9$_{\mathrm{NSF}}$
Flag–DHFR–Ub

Figure 1 Enhancement of Cup9 degradation by dipeptides with destabilizing N-terminal residues. **a**, The fusion protein used for pulse-chase analysis. The stable Flag–DHFR–ubiquitin reference portion of the fusion is co-translationally cleaved from Cup9$_{\mathrm{NSF}}$ by ubiquitin-specific protease (UBP). **b**, Pulse-chase analysis of Flag–DHFR–ubiquitin–Cup9$_{\mathrm{NSF}}$ in the presence of various dipeptides at a concentration of 10 mM. Dipeptides with either basic (Arg–Ala) or bulky hydrophobic (Leu–Ala) destabilizing N-terminal residues strongly enhance Cup9$_{\mathrm{NSF}}$ degradation, but only in strains expressing Ubr1. Dipeptides bearing a stabilizing N-terminal residue (Ala–Arg and Ala–Leu) do not alter Cup9$_{\mathrm{NSF}}$ degradation. **c**, Effects of different concentrations of Trp–Ala on the enhancement of Cup9$_{\mathrm{NSF}}$ degradation in wild-type (*UBR1*) cells. Dash indicates that pulse-chase analysis was performed in the absence of added dipeptides. Enhancement of Cup9$_{\mathrm{NSF}}$ degradation was detectable at 1 μM Trp-Ala Ub$^{\mathrm{R48}}$, ubiquitin containing mutation of Lys 48 to Arg (ref. 20).

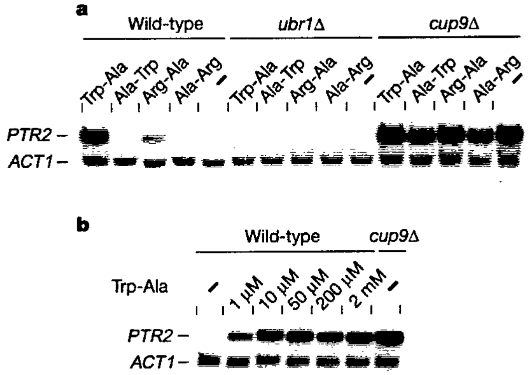

a
Wild-type
ubr1Δ
cup9Δ
Trp-Ala
Ala-Trp
Arg-Ala
Ala-Arg
PTR2 –
ACT1 –
b
Wild-type
cup9Δ
Trp-Ala
1 μM
10 μM
50 μM
200 μM
2 mM
PTR2 –
ACT1 –

Figure 2 Effects of dipeptides on expression of the dipeptide transporter gene *PTR2*. **a**, Induction of *PTR2* expression by dipeptides bearing destabilizing N-terminal residues (Trp–Ala and Arg–Ala) required both *UBR1* and *CUP9*. Dipeptides bearing a stabilizing N-terminal residue (Ala–Trp and Ala–Arg) had no effect on *PTR2* expression. *PTR2* mRNA and the *ACT1* mRNA loading control are indicated. **b**, Effect of different concentrations of Trp–Ala on the levels of *PTR2* mRNA.

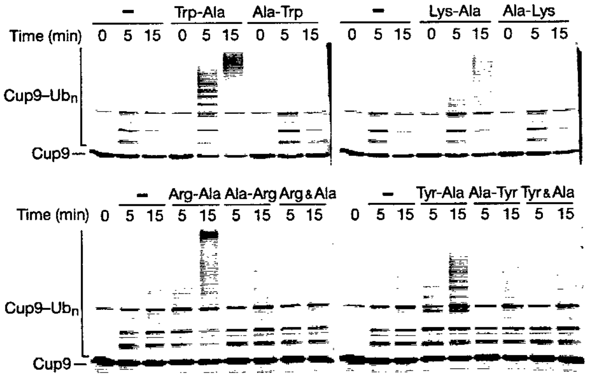

–
Trp-Ala
Ala-Trp
–
Lys-Ala
Ala-Lys
Time (min) 0 5 15 0 5 15 0 5 15 0 5 15 0 5 15 0 5 15
Cup9–Ub$_n$
Cup9 –
–
Arg-Ala
Ala-Arg
Arg & Ala
–
Tyr-Ala
Ala-Tyr
Tyr & Ala
Time (min) 0 5 15 5 15 5 15 5 15 0 5 15 5 15 5 15 5 15
Cup9–Ub$_n$
Cup9 –

Figure 3 *In vitro* ubiquitylation of Cup9 is enhanced by dipeptides bearing destabilizing N-terminal residues. Reactions consisting of purified Uba1 (E1), Rad6 (E2), Ubr1 (E3), ubiquitin, ATP and radiolabelled Cup9 were supplemented with the indicated dipeptides or amino acids (top panel, 2 μM, bottom panel, 10 μM), or left unsupplemented (–), and allowed to proceed at 30 °C for the designated times. Radiolabelled input Cup9, and its multi-ubiquitylated derivatives are indicated. Ub, ubiquitin.

gene[5]. Thus, the dipeptide-induced acceleration of Cup9 degradation would be expected to increase the levels of *PTR2* messenger RNA, ultimately leading to an increase in dipeptide uptake. We tested this by examining the levels of *PTR2* mRNA in the presence or absence of dipeptides in the medium. At 25 μM, both Trp–Ala and Arg–Ala induced *PTR2* expression in the wild-type (*UBR1*) strain (Fig. 2a). Both Ubr1 and Cup9 were required for these effects, as the expression of *PTR2* was not altered by dipeptides in *ubr1Δ* and *cup9Δ* strains. Induction of *PTR2* mRNA could be observed at 1 μM Trp–Ala, and increased at higher dipeptide concentrations (Fig. 2b), in agreement with the observed changes in the half-life of Cup9 (Fig. 1c).

A plausible mechanism of the enhancement effect is that a dipeptide interacts with either the basic or the hydrophobic N-terminal residue-binding sites of Ubr1, and a distinct (third) substrate-binding site of Ubr1 recognizes the internal degron of Cup9. In this model, the interaction of Ubr1 with dipeptides allosterically increases the ability of the Ubr1–Rad6 (E3–E2) complex to ubiquitylate Cup9. To test whether dipeptides act directly through Ubr1, we examined the effect of dipeptides on Cup9 ubiquitylation in an *in vitro* system consisting of the following purified components: Ubr1 (E3), Rad6 (E2), Uba1 (E1), ubiquitin, ATP and radiolabelled Cup9. Cup9 was significantly multi-ubiquitylated, in a Ubr1/Rad6-dependent reaction (data not shown), in the absence of added dipeptides (Fig. 3). This result was consistent with the relatively rapid *in vivo* degradation of Cup9p ($t_{1/2}$ ~5 min) in the absence of dipeptides (Fig. 1c).

The addition of dipeptides bearing either type of destabilizing N-terminal residue to the *in vitro* system substantially stimulated the Ubr1-dependent multi-ubiquitylation of Cup9 (Fig. 3). Dipeptides of the same composition but bearing a stabilizing N-terminal residue did not stimulate multi-ubiquitylation, nor did the amino-acid components of these dipeptides (Fig. 3). These results demonstrate that dipeptides act directly through Ubr1, without an intermediate signalling pathway. The underlying allosteric mechanism may involve increased affinity of Ubr1 for Cup9, or enhanced ubiquitylation activity of the Ubr1–Rad6 complex towards Cup9, or both.

This work shows that the two binding sites of Ubr1 that interact with destabilizing N-terminal residues can act as allosteric sites that enable Ubr1 to sense the presence of imported dipeptides and to accelerate degradation of the Cup9 repressor, resulting in an appropriate induction of the Ptr2 transporter (Fig. 4a–d). This model predicts that a dipeptide bearing a destabilizing N-terminal residue, for example Leu–Ala, should stimulate its own uptake, in contrast to Ala–Leu. This prediction was borne out when we tested the ability of these two leucine-containing dipeptides to support the growth of *S. cerevisiae* auxotrophic for leucine (Fig. 4e).

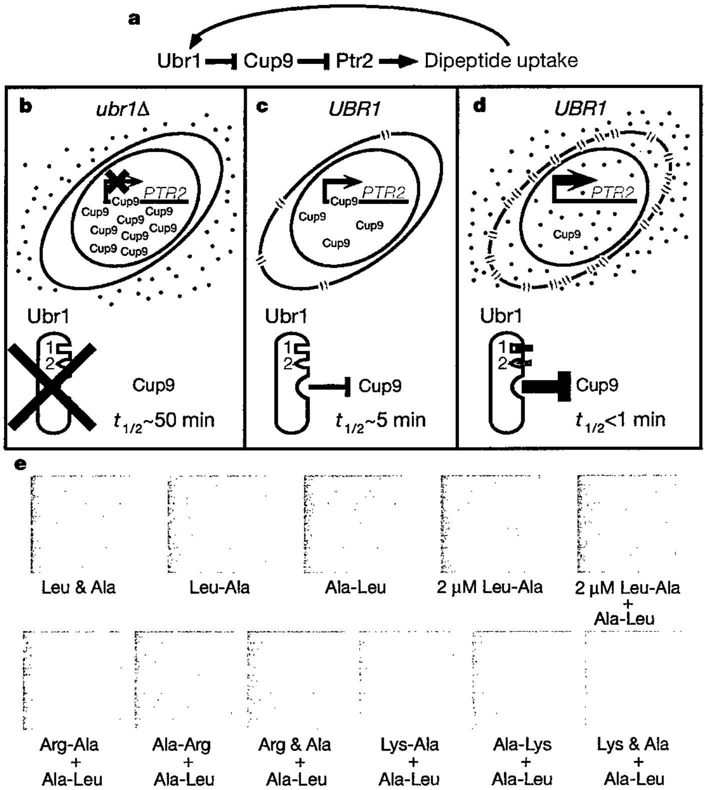


Figure 4 Feedback regulation of peptide import in *S. cerevisiae*. **a**, The peptide transport circuit. **b**, Ubr1 is required for dipeptide uptake. In the absence of Ubr1 (*ubr1Δ*), the transcriptional repressor Cup9 is long-lived, accumulates to high levels and extinguishes expression of the *PTR2* gene. Thus, *ubr1Δ* cells cannot import dipeptides (red dots). **c**, In a wild-type (*UBR1*) cell growing in the absence of extracellular dipeptides, Ubr1 targets Cup9 for degradation ($t_{1/2}$ ~5 min), resulting in a moderate concentration of Cup9 and weak but significant expression of the Ptr2 transporter (blue double ovals). **d**, In wild-type (*UBR1*) cells growing in the presence of extracellular dipeptides, some of which bear destabilizing N-terminal residues, the imported dipeptides bind to either the basic (type 1; red) or the hydrophobic (type 2; green) residue-binding site of Ubr1. Binding of either type of dipeptide to Ubr1 allosterically increases the rate of Ubr1-mediated degradation of Cup9. Peptides of both types are shown bound to Ubr1, but the binding of either peptide accelerates Cup9 degradation. The resulting decrease of the half-life of Cup9 from ~5 min to <1 min results in a very low concentration of Cup9, and consequently to strong induction of the Ptr2 transporter. **e**, Colony formation assays. An *S. cerevisiae* strain requiring leucine for growth was incubated on plates supplemented with either dipeptides or their amino-acid constituents at the following concentrations: Leu–Ala, 230 μM (2 μM where indicated); Ala–Leu, 230 μM; Arg–Ala, 10 μM; Ala–Arg, 10 μM; Lys–Ala, 1 μM; Ala–Lys, 1 μM; Arg & Ala, 10 μM each, Lys & Ala, 1 μM each.

Food sources that *S. cerevisiae* encounters outside the laboratory setting are likely to contain mixtures of short peptides, a subset of which would be capable of activating the Ubr1-based positive feedback circuit. We modelled this situation by providing cells with a mixture of Leu–Ala at a low concentration (2 μM) and Ala–Leu at a high concentration (230 μM). Although neither dipeptide supplement alone could satisfy the strain's requirement for leucine, a mixture of the two dipeptides supported robust growth (Fig. 4e). Moreover, a mixture of Ala–Leu (230 μM) and a dipeptide lacking leucine but bearing a destabilizing N-terminal residue (10 μM Arg–Ala or 1 μM Lys–Ala) also rescued growth. In contrast, dipeptides bearing stabilizing N-terminal residues, or the amino-acid components of these dipeptides, could not rescue the growth of Leu$^-$ cells in the presence of 230 μM Ala–Leu (Fig. 4e).

This work establishes for the first time that the activity of an E3 can be directly linked to the presence of an environmental signal through an allosteric interaction with a small compound. Specifically, dipeptides bearing destabilizing N-terminal residues are shown to act as allosteric activators of Ubr1, enhancing its ability to support the ubiquitylation and degradation of Cup9. Physiologically, this results in a positive feedback circuit governing the uptake of peptides (Fig. 4). Through their binding to Ubr1, the imported dipeptides accelerate degradation of Cup9, thereby derepressing the synthesis of the Ptr2 transporter and enhancing the cell's ability to import di- and tripeptides. As most cells have the capacity to import peptides, the results of this work suggest that peptide import may be regulated similarly by Ubr1 homologues in metazoans[23] and by the ClpAP-dependent N-end rule pathway in *Escherichia coli*[24]. The ubiquitin system is either known or suspected to be important in the control of intermediary metabolism and the transport of small molecules across membranes[2,3]. Our findings suggest that these compounds, or their enzymatically produced derivatives, may modulate the functions of E3s in the ubiquitin system similarly to the effects observed here with dipeptides and Ubr1. ☐

Methods

Yeast strains and plasmids

The *S. cerevisiae* strains used in pulse-chase experiments, JD52 (*MATα lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-3,112*), JD55 (*ubr1Δ*::HIS3), have been described[25]. Flag–DHFR–ubiquitin–Cup9$_{NSF}$ was expressed from the P$_{MET25}$ promoter on the centromeric vector p416MET25 (ref. 26). Construction details are available upon request. Strains used for northern analyses were AVY30 (*MATα leu2-3,112 ubr1Δ::LEU2*), AVY 31 (*MATα leu2-3,112 cup9Δ::LEU2*) and AVY32 (*MATα LEU2*), constructed in the RJD350 background (*MATα leu2-3,112*; a gift from R. Deshaies) using restriction fragments obtained from plasmids pSOB30 (ref. 6), pCB119 (ref. 5) and pJJ252 (ref. 27), respectively. The RJD350 strain was used for the colony formation assay (Fig. 4e). SHM plates[5] were supplemented with dipeptides or amino acids at the following concentrations: 230 μM Leu–Ala (or 2 μM where indicated), 230 μM Ala–Leu, 230 μM each of Leu and Ala, 10 μM Arg–Ala, 10 μM Ala–Arg, 1 μM Lys–Ala, 1 μM Ala–Lys, 10 μM each of Arg and Ala, or 1 μM each of Lys and Ala.

Pulse-chase analysis

Cells were cultured in SHM[5] with auxotrophic supplements. Dipeptides were added to cultures at an absorbance of ~0.6 at 600nm and incubation continued for 2.5 h in the experiments in Fig. 1b, and for 30 min for Fig. 1c. Cells were harvested, washed in 0.8 ml of SHM, resuspended in 0.4 ml of SHM, and labelled for 5 min at 30 °C with 0.16 mCi of ^{35}S-EXPRESS (New England Nuclear). Cells were pelleted and resuspended in fresh SHM containing 4 mM L-methionine and 2 mM L-cysteine. Samples (0.1ml) were taken at the time points indicated and transferred to chilled tubes, each containing 0.5 ml of 0.5-mm glass beads, 0.7 ml of ice-cold lysis buffer (1% Triton-X100, 0.15 M NaCl, 5 mM EDTA, 50 mM sodium HEPES, pH 7.5), and a mixture of protease inhibitors (final concentrations 1 mM phenylmethylsulphonyl fluoride, 2 μg ml^{-1} aprotinin, 0.5 μg ml^{-1} leupeptin and 0.7 μg ml^{-1} pepstatin). Extracts were prepared and immunoprecipitations carried out as described[25], using anti-Flag M2 resin (Sigma). Immunoprecipitates were fractionated by 13% SDS–polyacrylamide gel electrophoresis (PAGE), and detected by autoradiography.

RNA preparation and northern analysis

Cells were cultured in SHM to an absorbance of ~0.6 at 600 nm. The indicated dipeptides were then added to a 25 μM final concentration, and the incubation was continued for an additional 30 min. Total RNA was prepared[28] and 25 μg samples were electrophoresed in 1% formaldehyde–agarose gels, followed by blotting for northern analysis[29].

Ubr1-dependent *in vitro* ubiquitylation system

The components of this system were purified as follows. N-terminally hexahistidine-tagged Uba1 was overexpressed in *S. cerevisiae* and purified by fractionation over Ni-NTA, ubiquitin affinity and Superdex-200 columns. Rad6 was overexpressed in *E. coli* and purified by fractionation over DEAE, Mono-Q and Superdex-75 columns. N-terminally Flag-tagged Ubr1 was overexpressed in *S. cerevisiae*, and purified by fractionation over anti-Flag M2, Ubc2p affinity and Superdex-200 columns. N-terminally Flag-tagged, C-terminally hexahistidine-tagged Cup9 was expressed and radiolabelled in *E. coli*, and purified by fractionation over Ni-NTA and anti-Flag M2 columns. Details of these expression and purification protocols are available upon request.

The *in vitro* ubiquitylation reactions contained the following components: 7 μM ubiquitin, 50 nM Uba1, 50 nM Rad6, 50 nM Ubr1, 550 nM ^{35}S-labelled Cup9, 25 mM HEPES/KOH (pH 7.5), 25 mM KCl, 5 mM $MgCl_2$, 2 mM ATP, 0.1 mM dithiothreitol and 0.5 mg ml^{-1} ovalbumin as a carrier protein. Dipeptides or amino acids, as indicated, were added to a final concentration of 2 μM (top panel, Fig. 3) or 10 μM (bottom panel, Fig. 3). All components except Uba1 were mixed on ice for 10 min; Uba1 was then added and reactions shifted to 30 °C. After the indicated times, the reactions were terminated by adding an equal volume of two times SDS–PAGE loading buffer and heating at 95 °C for 5 min, followed by 8% SDS–PAGE.

Received 5 January; accepted 5 April 2000.

1. Laney, J. D. & Hochstrasser, M. Substrate targeting in the ubiquitin system. *Cell* **97,** 427–430 (1999).
2. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **76,** 425–479 (1998).
3. Varshavsky, A. The ubiquitin system. *Trends Biochem. Sci.* **22,** 383–387 (1997).
4. Baumeister, W., Walz, J., Zühl, F. & Seemüller, E. The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92,** 367–380 (1998).
5. Byrd, C., Turner, G. C. & Varshavsky, A. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. *EMBO J.* **17,** 269–277 (1998).
6. Bartel, B., Wünning, I. & Varshavsky, A. The recognition component of the N-end rule pathway. *EMBO J.* **9,** 3179–3189 (1990).
7. King, R. W., Deshaies, R. J., Peters, J. M. & Kirschner, M. W. How proteolysis drives the cell cycle. *Science* **274,** 1652–1659 (1996).
8. Kotani, S., Tanaka, H., Yasuda, H. & Todokoro, K. Regulation of APC activity by phosphorylation and regulatory factors. *J. Cell Biol.* **146,** 791–800 (1999).
9. Bachmair, A., Finley, D. & Varshavsky, A. *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* **234,** 179–186 (1986).
10. Bachmair, A. & Varshavsky, A. The degradation signal in a short-lived protein. *Cell* **56,** 1019–1032 (1989).
11. Varshavsky, A. The N-end rule: functions, mysteries, uses. *Proc. Natl Acad. Sci. USA* **93,** 12142–12149 (1996).
12. Xie, Y. M. & Varshavsky, A. The E2-E3 interaction in the N-end rule pathway: the RING-H2 finger of E3 is required for the synthesis of multiubiquitin chain. *EMBO J.* **18,** 6832–6844 (1999).
13. Dohmen, R. J., Madura, K., Bartel, B. & Varshavsky, A. The N-end rule is mediated by the UBC2(RAD6) ubiquitin-conjugating enzyme. *Proc. Natl Acad. Sci. USA* **88,** 7351–7355 (1991).
14. Reiss, Y., Kaim, D. & Hershko, A. Specificity of binding of N-terminal residues of proteins to ubiquitin-protein ligase. Use of amino acid derivatives to characterize specific binding sites. *J. Biol. Chem.* **263,** 2693–2269 (1988).
15. Gonda, D. K. *et al.* Universality and structure of the N-end rule. *J. Biol. Chem.* **264,** 16700–16712 (1989).
16. Baker, R. T. & Varshavsky, A. Inhibition of the N-end rule pathway in living cells. *Proc. Natl Acad. Sci. USA* **87,** 2374–2378 (1991).
17. Alagramam, K., Naider, F. & Becker, J. M. A recognition component of the ubiquitin system is required for peptide transport in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **15,** 225–234 (1995).
18. Knight, S. A. B., Tamai, K. T., Kosman, D. J. & Thiele, D. J. Identification and analysis of a *Saccharomyces cerevisiae* copper homeostasis gene encoding a homeodomain protein. *Mol. Cell. Biol.* **14,** 7792–7804 (1994).
19. Wolberger, C., Vershon, A. K., Liu, B. S., Johnson, A. D. & Pabo, C. O. Crystal structure of a Matα-2 homeodomain–operator complex suggests a general model for homeodomain–DNA interactions. *Cell* **67,** 517–528 (1991).
20. Lévy, F., Johnsson, N., Rumenapf, T. & Varshavsky, A. Using ubiquitin to follow the metabolic fate of a protein. *Proc. Natl Acad. Sci. USA* **93,** 4907–4912 (1996).
21. Suzuki, T. & Varshavsky, A. Degradation signals in the lysine–asparagine sequence space. *EMBO J.* **18,** 6017–6026 (1999).
22. Barnes, D., Lai, W., Breslav, M., Naider, F. & Becker, J. M. PTR3, a novel gene mediating amino acid-inducible regulation of peptide transport in Saccharomyces cerevisiae. *Mol. Microbiol.* **29,** 297–310 (1998).
23. Kwon, Y. T. *et al.* The mouse and human genes encoding the recognition component of the N-end rule pathway. *Proc. Natl Acad. Sci. USA* **95,** 7898–7903 (1998).
24. Tobias, J. W., Shrader, T. E., Rocap, G. & Varshavsky, A. The N-end rule in bacteria. *Science* **254,** 1374–1377 (1991).
25. Ghislain, M., Dohmen, R. J., Levy, F. & Varshavsky, A. Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in *Saccharomyces cerevisiae*. *EMBO. J.* **15,** 4884–4899 (1996).
26. Mumberg, D., Muller, R. & Funk, M. Regulatable promoters of *Saccharomyces cerevisiae*—comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.* **22,** 5767–5768 (1994).
27. Jones, J. S. & Prakash, L. Yeast *Saccharomyces cerevisiae* selectable markers in pUC18 polylinkers. *Yeast* **6,** 363–366 (1990).
28. Schmitt, M. E., Brown, T. A. & Trumpower, B. L. A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **18,** 3091–3092 (1990).
29. Ausubel, F. M. *et al.* (eds) *Current Protocols in Molecular Biology.* (Wiley-Interscience, New York, 1996).

Acknowledgements

We thank A. Webster for her valuable contributions to establishing the *in vitro* system, and R. Deshaies, J. Dohmen, L. Prakash, H. Rao and J. Sheng for their gifts of plasmids and strains. We also thank C. Byrd, H. Rao, T. Iverson and especially R. Deshaies for helpful discussions and comments on the manuscript. This work was supported by grants to A.V. from the NIH.

Correspondence and requests for materials should be addressed to A.V.
(e-mail: avarsh@caltech.edu).

Physical association of ubiquitin ligases and the 26S proteasome

Youming Xie and Alexander Varshavsky[*]

Division of Biology, California Institute of Technology, Pasadena, CA 91125

Contributed by Alexander Varshavsky, January 19, 2000

The ubiquitin (Ub) system recognizes degradation signals of the target proteins through the E3 components of E3-E2 Ub ligases. A targeted substrate bears a covalently linked multi-Ub chain and is degraded by the ATP-dependent 26S proteasome, which consists of the 20S core protease and two 19S particles. The latter mediate the binding and unfolding of a substrate protein before its transfer to the interior of the 20S core. It is unclear how a targeted substrate is delivered to the 26S proteasome, inasmuch as Rpn10p, the only known proteasomal subunit that binds multi-Ub chains, has been found to be not essential for degradation of many proteins in the yeast *Saccharomyces cerevisiae*. Here we show that Ubr1p and Ufd4p, the E3 components of two distinct Ub ligases, directly interact with the 26S proteasome. Specifically, Ubr1p is shown to bind to the Rpn2p, Rpt1p, and Rpt6p proteins of the 19S particle, and Ufd4p is shown to bind to Rpt6p. These and related results suggest that a substrate-bound Ub ligase participates in the delivery of substrates to the proteasome, because of affinity between the ligase's E3 component and specific proteins of the 19S particle.

E3 | N-end rule | ubiquitin fusion degradation pathway | Ubr1p | Ufd4p

Regulated proteolysis by the ubiquitin (Ub) system plays essential roles in the cell cycle, differentiation, stress responses, and many other processes (1–5). Ub is a 76-residue protein whose covalent conjugation to other proteins, usually in the form of a multi-Ub chain, marks these proteins for processive degradation by the 26S proteasome, an ATP-dependent multisubunit protease (3, 6–8). The conjugation of Ub to other proteins involves the formation of a thioester between the C terminus of Ub and a specific cysteine of the Ub-activating (E1) enzyme. The Ub moiety of E1~Ub thioester thereafter is transesterified to a specific cysteine in one of several Ub-conjugating (E2) enzymes. The Ub moiety of E2~Ub thioester is conjugated, via the isopeptide bond, to the ε-amino group of either a substrate's Lys residue or a Lys residue of another Ub moiety, the latter reaction resulting in a substrate-linked multi-Ub chain (9–11).

Most E2 enzymes function in complexes with proteins called E3. The functions of E3s include the initial recognition of degradation signals (degrons) in the substrate proteins, with different E3s recognizing different classes of degrons. The E2-E3 complexes, referred to as Ub ligases, mediate the formation of substrate-linked multi-Ub chains (10, 12, 13). The ATP-dependent 26S proteasome, which processively degrades a targeted, ubiquitylated[†] substrate, consists of the 20S core protease and two 19S particles (6, 14). A 19S particle mediates the binding and unfolding of a substrate protein before its transfer to the interior of the 20S core (6, 15–19). It is unclear how a targeted substrate is delivered to the 26S proteasome, inasmuch as Rpn10p (Mcb1p/Sun1p), the only known proteasomal subunit that binds multi-Ub chains (20), is not essential for degradation of many proteins in the yeast *Saccharomyces cerevisiae* (21).

Here we show that Ubr1p and Ufd4p, the E3 components of two distinct Ub-dependent proteolytic pathways (13, 22, 23), directly interact with the 26S proteasome. Specifically, Ubr1p is shown to bind to the Rpn2p (Sen3p), Rpt1p (Cim5p), and Rpt6p (Cim3p/Sug1p) proteins of the 19S particle, and Ufd4p is shown to bind to Rpt6p. These results suggest a mechanism for the delivery of protein substrates to the proteasome.

Materials and Methods

Protein Expression in *S. cerevisiae* and *Escherichia coli*. Plasmid construction protocols are available on request. *S. cerevisiae RPN1* was isolated as a multicopy suppressor of the toxicity of co-overproduced Ubr1p and Rad6p (Ubc2p), which were expressed from the bidirectional $P_{GAL1,10}$ promoter as described (24). *RPN2*, *RPN10*, and *RPN12* with their own promoter regions, and the ORFs of *RPN3, RPN11, RPT1, RPT2, RPT6*, and *PRE6* were amplified by PCR from the DNA of *S. cerevisiae* YPH500. All PCR products were verified by DNA sequencing. For the toxicity suppression assays (Fig. 1), *RPN1, RPN2, RPN10*, and *RPN12* were subcloned into the high-copy vector pRS425 (25). The *RPT1* and *RPT6* ORFs were subcloned into the low-copy vector pRS314CUP1 derived from pRS314 (25), yielding p314CUP1RPT1 and p314CUP1RPT6, in which *RPT1* and *RPT6* were expressed from the P_{CUP1} promoter. For coimmunoprecipitation assays, the N-termini of Rpt6p, Pre6p, and Ufd4p were extended with the hemagglutinin (HA) epitope (23); alternatively, the FLAG epitope was added to the C termini of Rpt6p and Rpn1p (26). For glutathione *S*-transferase (GST)-pulldown assays, the ORFs of *RPN1, RPN2, RPN3, RPN10, RPN11, RPN12, RPT1, RPT2, RPT6*, and *PRE6* were fused in-frame to the 3′ end of GST-coding sequence in pGEX-4T-3 (Amersham Pharmacia). The C-terminally FLAG-tagged alleles of *RPN1* and *RPN10* were subcloned into pET-11c (Novagen). *E. coli* BL21 (DE3) (26) was used to express the GST fusions, as well as Rpn1p-FLAG and Rpn10p-FLAG.

Pulse–Chase and GST-Pulldown Assays. The Ub protein reference (UPR)-based plasmids expressing dihydrofolate reductase (DHFR)-HA-Ub^{R48}-X-βgal (β-galactosidase) (see *Results* for notations) were described (27). The pulse–chase procedures (Fig. 2) were described as well (22, 28). In GST-pulldown assays, $\approx$1 μg of a GST fusion protein or GST alone was diluted to 0.5 ml in the loading buffer (1% Triton X-100/10% glycerol/0.5 M NaCl/1 mM EDTA/50 mM Tris·HCl, pH 8.0), and incubated with 10 μl (bed volume) of the glutathione-agarose beads (Sigma) for 1 h at 4°C. The beads were washed three times with

---

Abbreviations: Ub, ubiquitin; HA, hemagglutinin; GST, glutathione *S*-transferase; UPR, Ub protein reference; DHFR, dihydrofolate reductase; βgal, β-galactosidase; UFD, Ub fusion degradation.

[*] To whom reprint requests should be addressed at: Division of Biology, 147-75, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125. E-mail: avarsh@caltech.edu.

[†] Ubiquitin whose C-terminal (Gly-76) carboxyl group is covalently linked to another compound is called the ubiquityl moiety, the derivative terms being ubiquitylation and ubiquitylated. The term Ub refers to both free ubiquitin and the ubiquityl moiety. This nomenclature (5), which also is recommended by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (41), brings ubiquitin-related terms in line with the standard chemical terminology.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.060025497.
Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.060025497

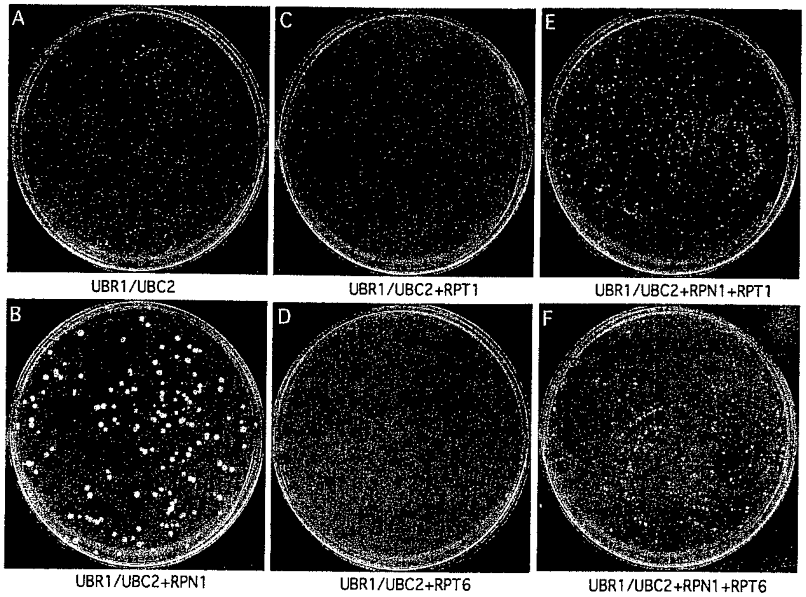


Fig. 1. Toxicity of overexpressed N-end rule pathway is decreased by overexpressed Rpn1p and enhanced by overexpressed Rpt1p or Rpt6p. The *S. cerevisiae* strain JD52 (23) was transformed with pKM1313, a low-copy, *URA3*-based plasmid expressing *UBR1* and *RAD6* from the bidirectional $P_{GAL1,10}$ promoter (24). The pKM1313-containing JD52 cells were transformed either with the pRS425 vector (*A*), *RPN1* (expressed from its own promoter in pRS425) (*B*), *RPT1* (expressed from the P_{CUP1} promoter in p314CUP1RPT1) (*C*), or *RPT6* (expressed from the P_{CUP1} promoter in p314CUP1RPT6) (*D*). (*E* and *F*) JD52 cells carrying both pKM1313 and the *RPN1*-expressing plasmid were further transformed with p314CUP1RPT1 (*E*) and p314CUP1RPT6 (*F*), respectively. Shown here are plates with cells grown in the galactose-containing minimal medium containing 0.2 mM $CuSO_4$ for 4 days at 30°C. On dextrose-containing plates (no overexpression of Ubr1p and Rad6p), all transformants grew at similar rates, irrespective of the presence of overexpressed Rpn1p, Rpt1p, or Rpt6p (data not shown).

1 ml of the binding buffer (0.05% Triton X-100/10% glycerol/50 mM NaCl/50 mM Na-Hepes, pH 7.5). The washed beads were incubated with cell extracts containing FLAG-Ubr1p (Fig. 3*A*), or with purified FLAG-Ubr1p (Fig. 3*B*), or with cell extracts containing either Rpn1p-FLAG or Rpn10p-FLAG (Fig. 4 *A* and *B*), or HA-Ufd4p (Fig. 5*A*) at 4°C for 2 h. The incubation buffer for the assay with purified FLAG-Ubr1p contained ovalbumin (Sigma) at 1 mg/ml in the binding buffer. The beads were washed three times with the binding buffer, followed by SDS/PAGE of the retained proteins and immunoblotting with either anti-FLAG antibody (Sigma) or anti-HA antibody (Babco, Berkeley, CA).

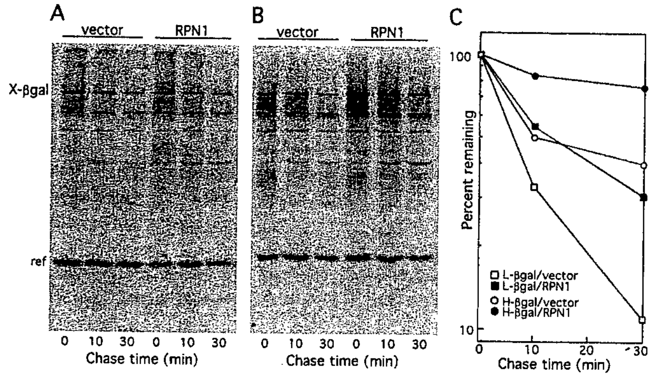


Fig. 2. Overexpression of Rpn1p inhibits degradation of N-end rule substrates. A UPR-based pulse–chase assay was used (see *Results*). *S. cerevisiae* JD52 (*UBR1*) cells coexpressing either DHFR-HA-Ub^{R48}-Leu-βgal (*A*) or DHFR-HA-Ub^{R48}-His-βgal (*B*), and either Rpn1p (expressed from its own promoter in pRS425) or pRS425 alone were labeled with [^{35}S]methionine for 5 min at 30°C, followed by a chase for 0, 10, and 30 min. Cell extracts were immunoprecipitated with both anti-βgal and anti-HA antibodies, followed by SDS/12% PAGE, autoradiography, and quantitation using PhosphorImager (Molecular Dynamics). (*C*) His-βgal and Leu-βgal decay curves calculated from the UPR-based data in *A* and *B* as described (28, 32). The bands of X-βgals and the reference protein DHFR-HA-Ub^{R48} ("ref") are indicated.

Coimmunoprecipitation/Immunoblotting. FLAG-Ubr1p, bearing the N-terminal FLAG epitope, was expressed from the P_{ADH1} promoter in the low-copy vector pRS315 (25). The N-terminally HA-tagged HA-Rpt6p, HA-Pre6p, and HA-Ufd4p and the C-terminally FLAG-tagged Rpt6p-FLAG were expressed from the induced P_{CUP1} promoter in the low-copy vector pRS314 (25). The C-terminally FLAG-tagged Rpn1p-FLAG was expressed from its own promoter in pRS425. *S. cerevisiae* was grown in synthetic-dextrose medium (26, 28) containing 0.2 mM $CuSO_4$ to OD_{600} ≈1.0. Cells were disrupted by vortexing with glass beads in 0.05% Triton X-100, 50 mM NaCl, 50 mM Na-Hepes (pH 7.5) containing several protease inhibitors (27).

Results

Ubr1p, the E3 of the N-End Rule Pathway, Binds to the Proteasome. Ubr1p (also called N-recognin), the 225K E3 of the *S. cerevisiae* N-end rule pathway, targets proteins that bear certain (destabilizing) N-terminal residues. Two substrate-binding sites of Ubr1p recognize two classes of destabilizing N-terminal residues: basic and bulky hydrophobic (13). Yet another substrate-binding site of Ubr1p targets proteins that bear internal (non-N-terminal) degrons (29). Similar, but distinct, versions of the N-end rule pathway are present in all organisms examined, from prokaryotes to fungi and mammals (13). Previous work has shown that overexpression of the pathway's Ub ligase, through co-overexpression of its E2 (Rad6p, also called Ubc2p) and E3 (Ubr1p) components, is toxic to *S. cerevisiae* (24).

In the present work, we found that *RPN1* (*NAS1/HRD2*) (30, 31), which encodes a ≈110 K protein of the proteasome's 19S particle, was a multicopy suppressor of the toxicity of co-overexpressed Ubr1p and Rad6p (Fig. 1 *A* and *B*, and data not shown). Because Rpn1p is essential for cell viability (30), it could be that the toxicity of co-overexpressed Ubr1p and Rad6p

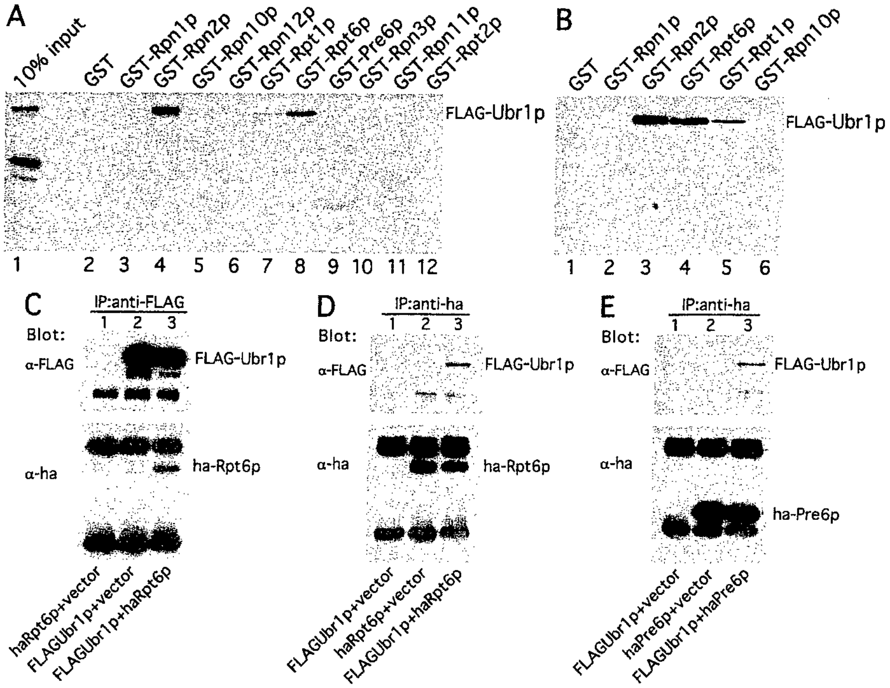

A
10% input
GST
GST-Rpn1p
GST-Rpn2p
GST-Rpn10p
GST-Rpn12p
GST-Rpt1p
GST-Rpt6p
GST-Pre6p
GST-Rpn3p
GST-Rpn11p
GST-Rpt2p
FLAG-Ubr1p
1 2 3 4 5 6 7 8 9 10 11 12
B
GST
GST-Rpn1p
GST-Rpn2p
GST-Rpt6p
GST-Rpt1p
GST-Rpn10p
FLAG-Ubr1p
1 2 3 4 5 6
C
IP:anti-FLAG
1 2 3
Blot:
α-FLAG
FLAG-Ubr1p
α-ha
ha-Rpt6p
haRpt6p+vector
FLAGUbr1p+vector
FLAGUbr1p+haRpt6p
D
IP:anti-ha
1 2 3
Blot:
α-FLAG
FLAG-Ubr1p
α-ha
ha-Rpt6p
FLAGUbr1p+vector
haRpt6p+vector
FLAGUbr1p+haRpt6p
E
IP:anti-ha
1 2 3
Blot:
α-FLAG
FLAG-Ubr1p
α-ha
ha-Pre6p
FLAGUbr1p+vector
haPre6p+vector
FLAGUbr1p+haPre6p

Fig. 3. Ubr1p, the E3 of the N-end rule pathway, is physically associated with the 26S proteasome. (*A* and *B*) Ubr1p interacts with Rpn2p, Rpt1p, and Rpt6p in GST-pulldown assays. Extracts of *S. cerevisiae* containing overexpressed FLAG-Ubr1p (*A*) or the purified FLAG-Ubr1p protein (*B*) were incubated with glutathione-agarose beads preloaded with the indicated GST fusions. The retained proteins were eluted, fractionated by SDS/8% PAGE, and immunoblotted with anti-FLAG antibody. Approximately equal amounts of different GST fusions were immobilized on glutathione-agarose beads in these assays, as verified by Coomassie staining (data not shown). (*C* and *D*) *In vivo* association of Ubr1p and Rpt6p. Extracts of *S. cerevisiae* AVY107 (*ubr1*Δ) expressing either both FLAG-Ubr1p and HA-Rpt6p, or FLAG-Ubr1p alone, or HA-Rpt6p alone were incubated with anti-FLAG antibody (*C*) or anti-HA antibody (*D*). The immunoprecipitated proteins were separated by SDS/12% PAGE and transferred to nitrocellulose membrane. The top halves of *C* and *D* show the results of immunoblotting with anti-FLAG antibody; the bottom halves show the analogous data with anti-HA antibody. (*E*) Coimmunoprecipitation of Pre6p and Ubr1p. Extracts of *S. cerevisiae* AVY107 (*ubr1*Δ) expressing both FLAG-Ubr1p and HA-Pre6p, FLAG-Ubr1p alone, or HA-Pre6p alone were incubated with anti-HA antibody, followed by the immunoprecipitation/immunoblotting described in *C* and *D*.

resulted in part from overdegradation of Rpn1p, which would have accounted for the observed alleviation of toxicity through overexpression of Rpn1p. However, pulse–chase assays with Rpn1p have shown it to be a long-lived protein in either wild-type cells or cells co-overexpressing Ubr1p and Rad6p (data not shown). It also was found that overexpression of Rpn1p had no effect on the expression levels of co-overexpressed Ubr1p and Rad6p (data not shown).

We then tested, using pulse–chase assays, whether overexpression of Rpn1p perturbed the degradation of normally short-lived substrates of the N-end rule pathway. In these assays, the *E. coli* βgal-based test proteins were expressed as components of fusions of the form DHFR-HA-Ub^{R48}-X-βgal, where X was either His, a type 1 (basic) destabilizing residue, or Leu, a type 2 (hydrophobic) destabilizing residue (13); DHFR-HA was the C-terminally HA-tagged mouse DHFR; Ub^{R48} was the Ub moiety containing Arg instead of Lys at position 48 (27, 28, 32); and βgal was *E. coli* βgal. Deubiquitylating enzymes (33) cotranslationally cleave these fusions at the Ub-βgal junction, yielding the long-lived DHFR-HA-Ub^{R48} reference protein and a test protein, either His-βgal or Leu-βgal. A reference protein improves the assay's accuracy by serving as a built-in control for variations in the expression levels, immunoprecipitation efficiency, and other sources of data scatter in a pulse–chase assay. The above method for producing a reference protein is called the UPR technique (28, 32). As shown in Fig. 2, the *in vivo* degradation of both His-βgal and Leu-βgal substrates of the N-end rule pathway was significantly impaired in the presence of overexpressed Rpn1p.

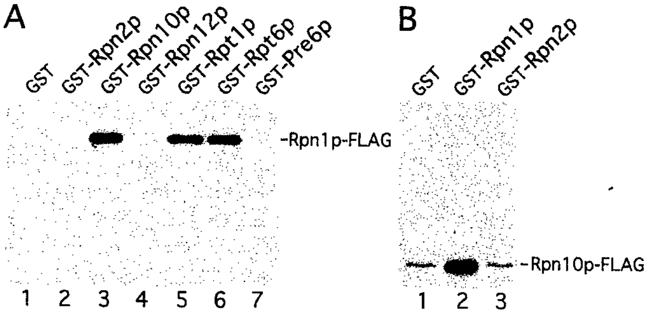

A
GST
GST-Rpn2p
GST-Rpn10p
GST-Rpn12p
GST-Rpt1p
GST-Rpt6p
GST-Pre6p
-Rpn1p-FLAG
1 2 3 4 5 6 7
B
GST
GST-Rpn1p
GST-Rpn2p
-Rpn10p-FLAG
1 2 3

Fig. 4. Rpn1p interacts with Rpt1p, Rpt6p, and Rpn10p. *E. coli*-expressed Rpn1p-FLAG (*A*) and Rpn10p-FLAG (*B*), both carrying the FLAG epitope at the C terminus, were incubated with glutathione-agarose beads preloaded with the indicated GST fusions. The retained proteins were fractionated by SDS/PAGE, followed by immunoblotting with anti-FLAG antibody. Approximately equal amounts of different GST fusions were immobilized on glutathione-agarose beads in these assays, as verified by Coomassie staining (data not shown).

To determine whether other proteins of the 19S particle were similar to Rpn1p in its ability to suppress the toxicity of overexpressed N-end rule pathway, we overexpressed Rpn2p, Rpn10p, and Rpn12p, by transforming cells that co-overexpressed Ubr1p and Rad6p (from the $P_{GAL1,10}$ promoter) with a high-copy vector expressing, separately, Rpn2p, Rpn10p, and Rpn12p from their own promoters. We also tested Rpt1p and Rpt6p, two of the six ATPases of the 19S particle (16), by overexpressing them from the copper-induced P_{CUP1} promoter in a low-copy vector. In contrast to the results with Rpn1p (Fig. 1 *A* and *B*), none of these proteins could alleviate, upon

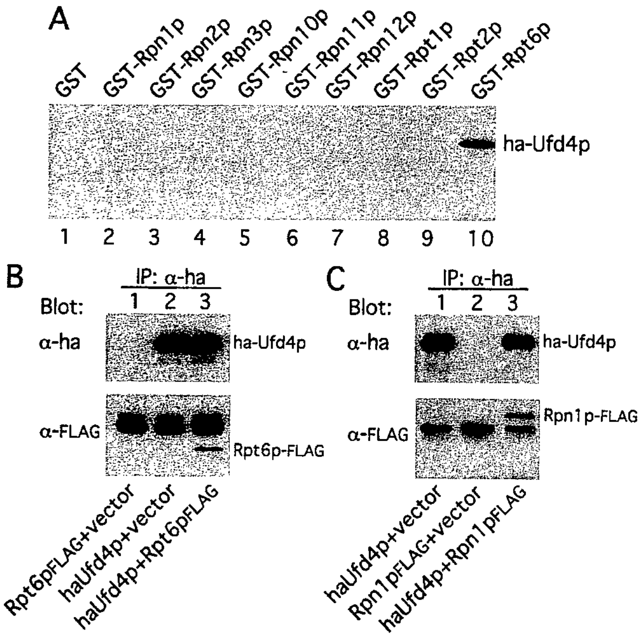


Fig. 5. Ufd4p, the E3 of the UFD pathway, is physically associated with the proteasome. (*A*) Ufd4p interacts with Rpt6p in the GST-pulldown assay. Extracts of *S. cerevisiae* containing overexpressed HA-Ufd4p were incubated with glutathione-agarose beads preloaded with different GST fusions, as indicated. The retained proteins were eluted, fractionated by SDS/8% PAGE, and immunoblotted with anti-HA antibody. Approximately equal amounts of different GST fusions were immobilized on glutathione-agarose beads in these assays, as verified by Coomassie staining (data not shown). (*B*) *In vivo* association of Ufd4p and Rpt6p. Extracts of *S. cerevisiae* JD52 (*UBR1*) expressing both HA-Ufd4p and Rpt6p-FLAG, HA-Ufd4p alone, or Rpt6p-FLAG alone were incubated with anti-HA antibody. The immunoprecipitated proteins were separated by SDS/10% PAGE and transferred to nitrocellulose membrane. (*Upper*) The results of immunoblotting with anti-HA antibody. (*Lower*) The data with anti-FLAG antibody. (*C*) Coimmunoprecipitation of Rpn1p and Ufd4p. Extracts of *S. cerevisiae* JD52 expressing both HA-Ufd4p and Rpn1p-FLAG, HA-Ufd4p alone, or Rpn1p-FLAG alone were incubated with anti-HA antibody, followed by the immunoprecipitation/immunoblotting procedure described in *B*, except that SDS/7% PAGE was used.

overexpression, the growth defect caused by co-overexpressed Ubr1p and Rad6p (data not shown). Moreover, we found that overexpression of either Rpt1p or Rpt6p had an effect opposite to that of overexpressed Rpn1p: the toxicity of co-overexpressed Ubr1p and Rad6p was higher in the presence of overexpressed Rpt1p or Rpt6p (Fig. 1 *C* and *D*; compare with Fig. 1 *A* and *B*). In a control experiment, the overexpression of Rpt1p or Rpt6p in cells that did not co-overexpress Ubr1p and Rad6p was not at all toxic (data not shown). These results suggested that Rpt1p and Rpt6p may interact with a component(s) of the N-end rule pathway, perhaps with Ubr1p, the pathway's E3. To verify this conjecture, we carried out GST-pulldown assays.

In these experiments, Ubr1p was N-terminally tagged with the FLAG epitope. The N-terminal FLAG tag did not impair the function of Ubr1p (A. Webster and A.V., unpublished data). Rpt1p and Rpt6p were expressed in *E. coli* as fusions to the C terminus of GST. Extracts of *S. cerevisiae* overexpressing FLAG-Ubr1p were incubated with glutathione-agarose beads preloaded with GST-Rpt1p, GST-Rpt6p, or GST alone. The bound proteins were eluted, fractionated by SDS/PAGE, and immunoblotted with a monoclonal anti-FLAG antibody. Remarkably, GST-Rpt6p was found to bind FLAG-Ubr1p; a smaller, but significant, amount of FLAG-Ubr1p also was bound by GST-Rpt1p, whereas no FLAG-Ubr1p was bound by GST alone (Fig. 3*A*, lanes 2, 7, and 8).

Could Ubr1p interact with other proteins of the 26S proteasome as well? Using the same assay, we discovered that GST-Rpn2p was also able to bind FLAG-Ubr1p, whereas all of the other tested proteasomal components (GST-Rpn1p, GST-Rpn3p, GST-Rpn10p, GST-Rpn11p, GST-Rpn12p, GST-Rpt2p, and GST-Pre6p, the latter a component of the 20S core proteasome) did not bind to FLAG-Ubr1p (Fig. 3*A*). Coomassie staining of the eluted, SDS/PAGE-fractionated proteins confirmed that the amounts of different GST fusions prebound to glutathione-agarose beads in the pulldown assays were approximately equal (data not shown).

To determine whether Ubr1p binds to Rpt1p, Rpt6p, and Rpn2p directly, we carried out pulldown assays with purified FLAG-Ubr1p (a gift from F. Du and A. Webster, California Institute of Technology) (Fig. 3*B*). FLAG-Ubr1p was overexpressed in *S. cerevisiae* and purified in three consecutive steps: anti-FLAG antibody affinity chromatography; Rad6p (E2) affinity chromatography; and gel filtration (F. Du, A. Webster, and A.V., unpublished data). In agreement with the results obtained by using extracts from cells overexpressing FLAG-Ubr1p (Fig. 3*A*), the purified FLAG-Ubr1p bound to the purified GST-Rpt1p, GST-Rpt6p, and GST-Rpn2p, but not to GST-Rpn1p, GST-Rpn10p, or GST alone (Fig. 3*B*). FLAG-Ubr1p used in this assay was purified from *S. cerevisiae* overexpressing this protein. Although FLAG-Ubr1p was at least 95% pure by Coomassie staining, one would like to purify *S. cerevisiae* Ubr1p from a noneukaryotic host as well, to preclude the unlikely possibility that a trace amount of a Ubr1p-interacting protein in the preparation of purified Ubr1p could mediate the binding of Ubr1p to Rpt1p, Rpt6p, and Rpn2p. Attempts to overexpress the 225K Ubr1p in *E. coli* have not been successful thus far.

Further tests used a coimmunoprecipitation/immunoblotting assay to examine the *in vivo* interaction between Ubr1p and Rpt6p. FLAG-Ubr1p was co-overexpressed in *ubr1*Δ *S. cerevisiae* with HA-Rpt6p, which bore the N-terminal HA-derived epitope. The controls included congenic cells expressing either FLAG-Ubr1p alone or HA-Rpt6p alone. Proteins were immunoprecipitated from cell extracts with anti-FLAG antibody, followed by SDS/PAGE and immunoblotting with either anti-HA antibody (Fig. 3*C Lower*) or anti-FLAG antibody (Fig. 3*C Upper*). The results indicated that HA-Rpt6p was specifically coprecipitated with FLAG-Ubr1p by anti-FLAG antibody. We also carried out the reciprocal coimmunoprecipitation assay and found that FLAG-Ubr1p was specifically coprecipitated with HA-Rpt6p by anti-HA antibody (Fig. 3*D*).

In the *in vitro* pulldown assays, FLAG-Ubr1p did not interact with GST-Pre6p, a GST fusion to a protein of the 20S core protease (Fig. 3*A*, lane 9). To determine whether Ubr1p is associated with the mature 26S proteasome *in vivo*, FLAG-Ubr1p and HA-Pre6p (N-terminally tagged with the HA epitope) were co-overexpressed in *ubr1*Δ cells. Proteins were precipitated from cell extracts with anti-HA antibody and separated by SDS/PAGE, followed by immunoblotting with either anti-FLAG or anti-HA antibody. FLAG-Ubr1p was found to be specifically coprecipitated with HA-Pre6p by anti-HA antibody (Fig. 3*E*). Thus, GST-pulldown assays with either crude or purified FLAG-Ubr1p (Fig. 3 *A* and *B*) and the coimmunoprecipitation assays (Fig. 3 *C*–*E*) demonstrated that Ubr1p, the E3 of the N-end rule pathway, is physically associated with the 26S proteasome. Moreover, Ubr1p was found to interact with more than one protein of the 19S particle (Fig. 3 *A* and *B*).

As described above, overexpression of Rpn1p, a non-ATPase component of the 19S particle, inhibited the activity of the N-end rule pathway (Fig. 2). What might be the mechanism of this inhibition? Overexpression of Rpn1p apparently did not have a significant effect on the function of the 26S proteasome, inasmuch as cells overexpressing Rpn1p were not hypersensitive to either an arginine analog canavanine or high temperature (data not shown). Hypersensitivity to these stressors is characteristic of yeast strains that have impaired proteasome components (6, 14). Overexpression of Rpn1p in wild-type cells (in the absence of co-overexpression of Ubr1p and Rad6p) did not result in a growth defect (data not shown). Furthermore, GST-pulldown assays indicated that GST-Rpn1p did not directly interact with Ubr1p (Fig. 3), ruling out the possibility that overproduced Rpn1p may decrease the toxicity of overexpressed N-end rule pathway by sequestering the overproduced Ubr1p. It was also unlikely that the overproduced, free Rpn1p competed with the 26S proteasome by interacting with substrate-linked multi-Ub chains, because Rpn1p was unable to bind isolated multi-Ub chains (data not shown).

As described above, overexpression of either Rpt1p or Rpt6p, the ATPase subunits of the 19S particle, was found to increase the toxicity of co-overexpressed Ubr1p and Rad6p (Fig. 1 *C* and *D*; compare with Fig. 1*A*). Therefore, we examined the possibility that the overexpressed Rpn1p may alleviate the toxicity of co-overexpressed Ubr1p and Rad6p by sequestering Rpt1p and Rpt6p, both of which were shown to interact with Ubr1p (Fig. 3). We began by determining, using the GST-pulldown assay, whether Rpn1p could directly bind to Rpt1p and/or Rpt6p. *S. cerevisiae* Rpn1p-FLAG (bearing the C-terminal FLAG epitope) was expressed in *E. coli* and incubated with glutathione agarose beads containing prebound GST-Rpt1p, GST-Rpt6p, or GST alone. The bound proteins were eluted, separated by SDS/PAGE, and immunoblotted with anti-FLAG antibody. Rpn1p was found to interact directly with both GST-Rpt1p and GST-Rpt6p, but not with GST alone (Fig. 4*A*, lanes 1, 5, and 6).

We also found that Rpn1p-FLAG directly interacted with GST-Rpn10p, the protein of the 19S particle that binds to multi-Ub chains (Fig. 4*A*). Rpn1p-FLAG did not interact, in this assay, with other tested proteins of the 19S particle (Rpn2p and Rpn12p) and also did not interact with the Pre6p protein of the 20S core (Fig. 4*A*). A direct Rpn1p-Rpn10p interaction also was observed in a reciprocal GST-pulldown assay, in which the *E. coli*-expressed Rpn10p-FLAG was shown to be retained by the prebound GST-Rpn1p, but not by GST-Rpn2p (Fig. 4*B*). Rpn1p, Rpn2p, Rpn10p, and the six Rpt ATPases Rpt1p-Rpt6p form the base of the 19S particle that is proximal to the 20S core of the 26S proteasome (16, 31).

Further tests indicated that overexpression of either Rpt1p or Rpt6p antagonized the ability of overexpressed Rpn1p to suppress the toxicity of overexpressed N-end rule pathway (Fig. 1 *E* and *F*; compare with Fig. 1*B*). Taken together, these results (Figs. 1, 3, and 4) strongly suggested that Rpn1p, which does not bind to Ubr1p, suppresses the toxicity of co-overexpressed Ubr1p and Rad6p at least in part through the demonstrated binding of Rpn1p to both Rpt1p and Rpt6p, either of which binds to Ubr1p. Schauber *et al.* (34) have shown that Rad23p, which bears an N-terminal Ub-like moiety, is another high-copy suppressor of the toxicity of overexpressed N-end rule pathway. Rad23p was found to interact with Rpt1p, Rpt6p, and Rpn10p (35, 36). In the present work, the same three proteins of the 19S particle were shown to interact with Rpn1p, another protein of this particle, suggesting that overexpression of Rpn1p and Rad23p suppresses the toxicity of overexpressed N-end rule pathway through similar mechanisms.

Ufd4p, the E3 of the Ub Fusion Degradation (UFD) Pathway, Binds to the Proteasome. The substrates of another Ub-dependent proteolytic system, termed the UFD pathway, include proteins bearing at their N termini a "nonremovable" Ub moiety (23, 37, 38). A partial or complete resistance of these Ub-containing proteins to deubiquitylating enzyme-mediated cleavage stems from either alterations of the last residue of Ub moiety or the presence of proline at the C-terminal side of the Ub-protein junction (23). Ufd4p, a member of the HECT family of E3 proteins (39), is the E3 of the *S. cerevisiae* UFD pathway (23). Sequence comparisons did not detect statistically significant similarities between the 225K Ubr1p and the 167K Ufd4p. We used the GST-pulldown and coimmunoprecipitation/immunoblotting assays to determine whether Ufd4p also could bind to the proteasome.

Ufd4p was N-terminally tagged with the HA epitope, which did not impair the function of Ufd4p (data not shown). In the GST-pulldown assay, extracts from *S. cerevisiae* overexpressing HA-Ufd4p were incubated with glutathione-agarose beads preloaded with GST-Rpn1p, GST-Rpn2p, GST-Rpn3p, GST-Rpn10p, GST-Rpn11p, GST-Rpn12p, GST-Rpt1p, GST-Rpt2p, GST-Rpt6p, or GST alone. The retained proteins were fractionated by SDS/PAGE and immunoblotted with anti-HA antibody. As shown in Fig. 5*A*, HA-Ufd4p was found to be specifically bound to GST-Rpt6p, one of three proteins of the 19S particle that interacted with Ubr1p, the E3 of the N-end rule pathway. None of the other tested proteasomal proteins were able to interact with HA-Ufd4p in the GST-pulldown assay (Fig. 5*A*).

We then examined the *in vivo* association of Ufd4p with Rpt6p. HA-Ufd4p was co-overexpressed in *S. cerevisiae* with Rpt6p-FLAG, which bore the C-terminal FLAG epitope. The controls included congenic cells expressing either HA-Ufd4p alone or Rpt6p-FLAG alone. Proteins were immunoprecipitated from cell extracts with anti-HA antibody, followed by SDS/PAGE and immunoblotting with either anti-HA antibody (Fig. 5*B Upper*) or anti-FLAG antibody (Fig. 5*B Lower*). The results indicated that Rpt6p-FLAG was specifically coprecipitated with HA-Ufd4p by anti-HA antibody.

If Ufd4p interacts with the mature proteasome, one would expect Rpn1p, which did not directly interact with Ufd4p (Fig. 5*A*), to be coimmunoprecipitated with HA-Ufd4p. To test this conjecture, HA-Ufd4p was co-overexpressed in *S. cerevisiae* with Rpn1p-FLAG, which bore the C-terminal FLAG epitope. The controls included congenic cells expressing either HA-Ufd4p alone or Rpn1p-FLAG alone. Extracts were immunoprecipitated with anti-HA antibody, followed by SDS/PAGE and immunoblotting with either anti-HA antibody (Fig. 5*C Upper*) or anti-FLAG antibody (Fig. 5*C Lower*). Rpn1p-FLAG was indeed specifically coprecipitated with HA-Ufd4p by anti-HA antibody. Taken together, the GST-pulldown and coimmunoprecipitation results (Fig. 5) demonstrated that Ufd4p, the E3 of the UFD pathway, was physically associated with the proteasome.

Discussion

It has been unclear how the Ub system delivers its substrates to the 26S proteasome for processive degradation. Two structural elements of a targeted, Ub-conjugated substrate could serve as ligands for the substrate's docking at the 26S proteasome: a substrate-linked multi-Ub chain and/or a cognate E3-E2 complex reversibly bound to the substrate's degron. Thus far, Rpn10p is the only proteasomal component known to bind multi-Ub chains; it also binds to chain-bearing model substrates (refs. 10, 20, and 21, and our unpublished data). However, *S. cerevisiae* stains lacking Rpn10p were found to be impaired in the degradation of only some of the model substrates and did not exhibit aberrant phenotypes characteristic of proteolysis-impaired proteasome mutants (10, 21).

Our discovery that Ubr1p and Ufd4p, the E3 components of, respectively, the N-end rule pathway and the UFD pathway, directly interact with specific proteins of the 26S proteasome suggests a general mechanism for the delivery of targeted substrates to the proteasome. In this model, the E3 component of a Ub-dependent pathway binds (as an E3-E2 complex) to a cognate degron of a target protein and initiates, either sequentially or concurrently, two sets of processes: (*i*) the formation of a substrate-linked multi-Ub chain, through the activity of substrate-bound Ub ligase (E3-E2) complex, and (*ii*) the delivery of an E3-bound substrate to the chaperone-like proteins of the 26S proteasome, through interactions between the E3 and specific proteins of the 19S particle. A key assumption of this model is that the demonstrated physical affinity between an E3 and the proteasome is essential for the activity (at least for the normal level of activity) of an E3-mediated proteolytic pathway. In the case of the N-end rule pathway, this prediction can be tested, for example, through the mapping of a site(s) in the 225K Ubr1p that mediates its binding to Rpn2p, Rpt1p, and Rpt6p of the 19S particle. The site(s) thus identified then could be mutated in ways that leave intact the other functions of Ubr1p, such as its binding to substrates, to the Rad6p E2 enzyme, and the previously described, RING finger-dependent activity of Ubr1p in the Rad6p/Ubr1p-mediated formation of a substrate-linked multi-Ub chain (27). A similar approach could be used to test the model's assumption in regard to the demonstrated Ufd4p-Rpt6p interaction.

Specific functions of either the Ufd4p-Rpt6p interaction or the interactions of Ubr1p with Rpn2p, Rpt1p, and Rpt6p of the 19S particle are unknown, in part because it is largely unknown how the proteins of this particle contribute to the proteasome-mediated proteolysis. It has been shown that Rpt1p-Rpt6p, the six ATPases of the 19S particle, are not functionally redundant, and that they cooperate in preparing individual substrates for degradation by the 20S core of the proteasome (40). It is likely that the functional role of the Ubr1p-Rpn2p interaction is different from those of the Ubr1p-Rpt1p and Ubr1p-Rpt6p interactions, because overexpression of Rpn2p was found to have virtually no effect on the toxicity of co-overexpressed Ubr1p and Rad6p, whereas overexpression of Rpt1p and Rpt6p increased this toxicity (see above). One possibility is that the non-ATPase Rpn2p protein of the 19S particle functions as the main Ubr1p affinity anchor, whereas the Ubr1p-binding ATPases Rpt1p and Rpt6p mediate the unfolding of a Ubr1p-bound substrate before its translocation into the 20S core. If so, the observed increase in the toxicity of overexpressed N-end rule pathway in the presence of free (overproduced) Rpt1p or Rpt6p may result from their binding to the Ubr1p moiety of the Ubr1p-Rad6p complex and their facilitation of the unfolding of Ubr1p-bound substrates before docking at the proteasome. Given the differences between Ubr1p and Ufd4p E3s in regard to their ligands in the 19S particle (unlike Ubr1p, the Ufd4p E3 did not bind to either Rpt1p or Rpn2p), it should be of interest to determine whether Rpt6p is, in fact, the only significant ligand of Ufd4p in the proteasome.

Several aspects of the proposed delivery mechanism remain unconstrained by the available evidence. For example, if the demonstrated Ubr1p-proteasome and Ufd4p-proteasome interactions (Figs. 3–5) prove relevant to the functional activity of the N-end rule and UFD pathways, one would like to determine how many of the distinct E3s in a cell interact with the proteasome and what proteins of the proteasome these E3s bind to. Another important question is whether the Ubr1p-dependent formation of a substrate-linked multi-Ub chain is required, *in vivo*, for the docking of a Ubr1p-bound substrate at the Ubr1-binding site of the proteasome. An alternative model is that these two Ubr1p-mediated, substrate-centered processes (the chain formation and the docking at the proteasome) take place concurrently and independently.

The previously proposed function of a substrate-linked multi-Ub chain is to serve as a dissociation-slowing device (9). Specifically, if the rate-limiting step that precedes the first proteolytic cleavages of a proteasome-bound substrate is the substrate's unfolding by chaperones of the 19S particle, then a decrease in the rate of dissociation of the proteasome-substrate complex, brought about by the multi-Ub chain, should facilitate substrate's degradation: the longer the allowed "waiting" time, the greater the probability of a required unfolding event (13). Similar considerations may apply to the function of E3-mediated binding of a substrate to the proteasome.

The main result of this work is that the E3 components of two distinct Ub ligases are physically associated with specific proteins of the 26S proteasome. We discussed the mechanistic and functional implications of this advance.

We thank F. Du and A. Webster for the gifts of purified FLAG-Ubr1p and the UPR-based plasmids, C. Pickart for isolated multi-Ub chains, H. Rao for the strain AVY107, and E. Johnson for the pRS313UFD4 plasmid. This work was supported by grants to A.V. from the National Institutes of Health (GM31530 and DK39520). Y.X. was supported in part by a postdoctoral fellowship from the National Institutes of Health.

1. Maniatis, T. (1999) *Genes Dev.* **13,** 505–510.
2. Nasmyth, K. (1999) *Trends Biochem. Sci.* **24,** 98–104.
3. Hershko, A. & Ciechanover, A. (1998) *Annu. Rev. Biochem.* **76,** 425–479.
4. Peters, J.-M., King, R. W. & Deshaies, R. J. (1998) in *Ubiquitin and the Biology of the Cell*, eds. Peters, J.-M., Harris, J. R. & Finley, D. (Plenum, New York), pp. 345–387.
5. Varshavsky, A. (1997) *Trends Biochem. Sci.* **22,** 383–387.
6. Baumeister, W., Walz, J., Zühl, F. & Seemüller, E. (1998) *Cell* **92,** 367–380.
7. Scheffner, M., Smith, S. & Jentsch, S. (1998) in *Ubiquitin and the Biology of the Cell*, eds. Peters, J.-M., Harris, J. R. & Finley, D. (Plenum, New York), pp. 65–98.
8. Hochstrasser, M. (1996) *Annu. Rev. Genet.* **30,** 405–439.
9. Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K. & Varshavsky, A. (1989) *Science* **243,** 1576–1583.
10. Pickart, C. M. (1997) *FASEB J.* **11,** 1055–1066.
11. Haas, A. J. & Siepman, T. J. (1997) *FASEB J.* **11,** 1257–1268.
12. Laney, J. D. & Hochstrasser, M. (1999) *Cell* **97,** 427–430.
13. Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 12142–12149.
14. Rechsteiner, M. (1998) in *Ubiquitin and the Biology of the Cell*, eds. Peters, J. M., Harris, J. R. & Finley, D. (Plenum, New York), pp. 147–189.
15. DeMartino, G. N. & Slaughter, C. A. (1999) *J. Biol. Chem.* **274,** 22123–22126.
16. Glickman, M. H., Rubin, D. M., Coux, O., Wefes, I., Pfeifer, G., Cjeka, Z., Baumeister, W., Fried, V. A. & Finley, D. (1998) *Cell* **94,** 615–623.
17. Rivett, A. J., Mason, G. G., Murray, R. Z. & Reidlinger, J. (1997) *Mol. Biol. Rep.* **24,** 99–102.
18. Coux, O., Tanaka, K. & Goldberg, A. L. (1996) *Annu. Rev. Biochem.* **65,** 801–817.
19. Hilt, W. & Wolf, D. H. (1996) *Trends. Biochem. Sci.* **21,** 96–102.
20. Deveraux, Q., Ustrell, V., Pickart, C. & Rechsteiner, M. (1994) *J. Biol. Chem.* **269,** 7059–7061.
21. van Nocker, S., Sadis, S., Rubin, D. M., Glickman, M., Fu, H., Coux, O., Wefes, I., Finley, D. & Vierstra, R. D. (1996) *Mol. Cell. Biol.* **16,** 6020–6028.
22. Bartel, B., Wünning, I. & Varshavsky, A. (1990) *EMBO J.* **9,** 3179–3189.
23. Johnson, E. S., Ma, P. C., Ota, I. M. & Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 17442–17456.
24. Madura, K. & Varshavsky, A. (1994) *Science* **265,** 1454–1458.
25. Sikorski, R. S. & Hieter, P. (1989) *Genetics* **122,** 19–27.
26. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G. & Struhl, K. (1998) *Current Protocols in Molecular Biology* (Wiley, New York).
27. Xie, Y. & Varshavsky, A. (1999) *EMBO J.* **18,** 6832–6844.
28. Suzuki, T. & Varshavsky, A. (1999) *EMBO J.* **18,** 6017–6026.
29. Byrd, C., Turner, G. C. & Varshavsky, A. (1998) *EMBO J.* **17,** 269–277.
30. Glickman, M. H., Rubin, D. M., Fried, V. A. & Finley, D. (1998) *Mol. Cell. Biol.* **18,** 3149–3162.
31. Finley, D., Tanaka, K., Mann, C., Feldmann, H., Hochstrasser, M., Vierstra, R., Johnston, S., Hampton, R., Haber, J., McCusker, J., *et al.* (1998) *Trends Biochem. Sci.* **23,** 244–245.
32. Lévy, F., Johnsson, N., Rumenapf, T. & Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 4907–4912.
33. Wilkinson, K. & Hochstrasser, M. (1998) in *Ubiquitin and the Biology of the Cell*, eds. Peters, J.-M., Harris, J. R. & Finley, D. (Plenum, New York), pp. 99–125.
34. Schauber, C., Chen, L., Tongaonkar, P., Vega, I., Lambertson, D., Potts, W. & Madura, K. (1998) *Nature (London)* **391,** 715–718.
35. Lambertson, D., Chen, L. & Madura, K. (1999) *Genetics* **153,** 69–79.
36. Hiyama, H., Yokoi, M., Masutani, C., Sugasawa, K., Maekawa, T., Tanaka, K., Hoeijmakers, J. H. J. & Hanaoka, F. (1999) *J. Biol. Chem.* **274,** 28019–28025.
37. Johnson, E. S., Bartel, B., W. & Varshavsky, A. (1992) *EMBO J.* **11,** 497–505.
38. Koegl, M., Hoppe, T., Schlenker, S., Ulrich, H. D., Mayer, T. U. & Jentsch, S. (1999) *Cell* **96,** 635–644.
39. Scheffner, M., Huibregtse, J. M., Vierstra, R. D. & Howley, P. M. (1993) *Cell* **75,** 495–505.
40. Rubin, D. M., Glickman, M., Larsen, C. N., Dhruvakumar, S. & Finley, D. (1998) *EMBO J.* **17,** 4909–4919.
41. Webb, E. C. (1992) *Enzyme Nomenclature* (Academic, New York), p. 527.

Science Reprint

Detecting and Measuring Cotranslational Protein Degradation in Vivo

Glenn C. Turner and Alexander Varshavsky*

22 September 2000, Volume 289, pp. 2117–2120

Copyright © 2000 by the American Association for the Advancement of Science

Detecting and Measuring Cotranslational Protein Degradation in Vivo

Glenn C. Turner and Alexander Varshavsky*

Nascent polypeptides emerging from the ribosome and not yet folded may at least transiently present degradation signals similar to those recognized by the ubiquitin system in misfolded proteins. The ubiquitin sandwich technique was used to detect and measure cotranslational protein degradation in living cells. More than 50 percent of nascent protein molecules bearing an amino-terminal degradation signal can be degraded cotranslationally, never reaching their mature size before their destruction by processive proteolysis. Thus, the folding of nascent proteins, including abnormal ones, may be in kinetic competition with pathways that target these proteins for degradation cotranslationally.

Nascent polypeptides emerging from the ribosome may, in the process of folding, present hydrophobic patches and other structural features that serve as degradation signals similar to those recognized by the ubiquitin (Ub) system in misfolded or otherwise damaged proteins (*1*). Whether a substantial fraction of nascent polypeptides is cotranslationally degraded is a long-standing question.

The Ub sandwich technique was developed to detect cotranslational protein degradation by measuring the steady-state ratio of two reporter proteins whose relative abundance is established cotranslationally. The technique requires that the polypeptide to be examined for cotranslational degradation, termed **B**, be sandwiched between two stable reporter domains, **A** and **C**, in a linear fusion protein (Fig. 1A). The three polypeptides are connected by Ub moieties, creating an **A**Ub-**B**Ub-**C**Ub fusion protein. Ub-specific processing proteases (UBPs) cotranslationally cleave such linear Ub fusions at the C-terminal residue of Ub (*2–4*), generating three independent polypeptides, **A**Ub, **B**Ub, and **C**Ub (*5*). UBP-mediated cleavage establishes a kinetic competition between two mutually exclusive events during the synthesis of **A**Ub-**B**Ub-**C**Ub: cotranslational UBP cleavage at the **B**Ub-**C**Ub junction to release the long-lived **C**Ub module or, alternatively, cotranslational degradation of the entire **B**Ub-**C**Ub nascent chain by the proteasome (*6*) (Fig. 1B). In the latter case, the processivity of proteasome-mediated degradation results in the destruction of the Ub moiety between **B** and **C** before it can be recognized by UBPs. The resulting drop in levels of the **C**Ub module relative to levels of **A**Ub, referred to as the C/A ratio, reflects the cotranslational degradation of domain **B** (Fig. 1B).

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA.

*To whom correspondence should be addressed. E-mail: avarsh@caltech.edu

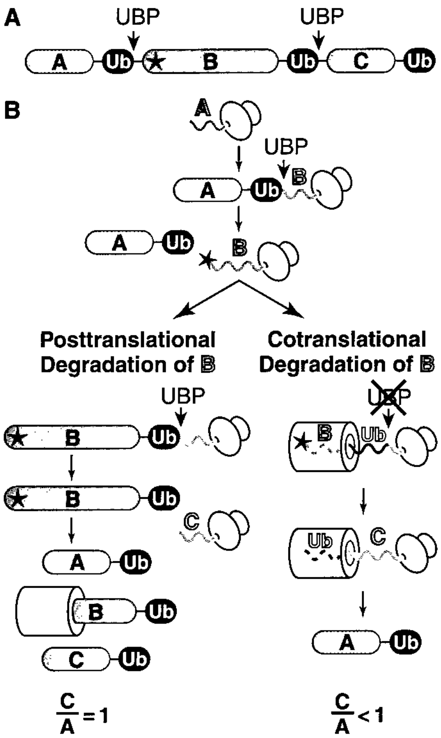

A
UBP
UBP
A
Ub
B
Ub
C
Ub
B
A
UBP
B
Posttranslational Degradation of B
Cotranslational Degradation of B
UBP
C/A = 1
C/A < 1

Fig. 1. The ubiquitin sandwich technique. (**A**) Organization of a Ub sandwich fusion. The polypeptide assayed for cotranslational degradation, **B**, is sandwiched between two stable reporter domains **A** and **C**. Red arrows indicate the locations of UBP cleavage sites. (**B**) The principle of the method. The reporter module **A**Ub is the first synthesized and is cotranslationally released from **B**, thereby providing a measure of the number of nascent **B** chains that initially emerge from the ribosome. If degradation of the emerging **B** domain, indicated by its insertion into the proteasome, is strictly posttranslational, UBP-mediated cleavage at the **B**Ub-**C** junction releases **C**Ub before **B** is degraded, so the molar yields of **C**Ub and **A**Ub are identical. However, if degradation of **B** can be cotranslational, a substantial fraction of **B**Ub-**C**Ub may be degraded as a unit. This will result in the molar yield of **C**Ub being lower than **A**Ub, the difference being a measure of cotranslational degradation.

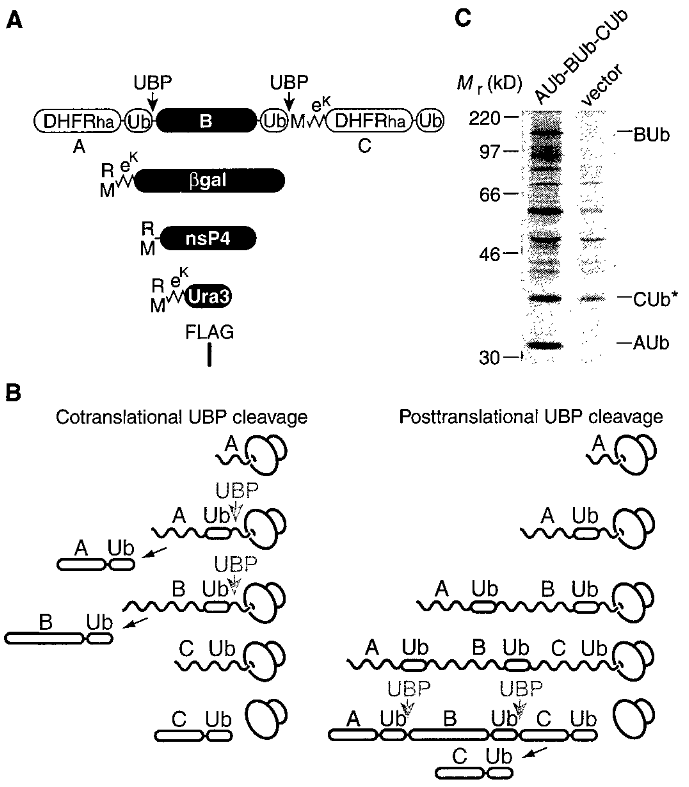


Fig. 2. UBP-mediated cleavage of ubiquitin sandwich fusions is cotranslational. (**A**) The protein fusions used. Domains **A** and **C** are mouse dihydrofolate reductase tagged with the influenza hemagglutinin-derived ha epitope (DHFRha). Domain **C** carries an N-terminal extension (e^K, see text), which makes it electrophoretically distinguishable from **A**. The different **B** domains are *E. coli* βGal, Sindbis virus RNA polymerase (nsP4), and *S. cerevisiae* Ura3p. Unstable versions of these domains have an N-terminal arginine (R) residue; stable versions have methionine (M). (**B**) The population of nascent chains produced by a radiolabeling pulse that is shorter than the time of translation of an **A**Ub-**B**Ub-**C**Ub fusion. Stretches of polypeptide containing radiolabel are in red; unlabeled stretches are in black. If UBPs efficiently cleave the nascent chain, free radiolabeled **A**Ub, **B**Ub, and **C**Ub should all be detected. If UBPs can cleave solely the full-length mature protein, only labeled **C**Ub will be observed. (**C**) The UBP cleavage of Ub sandwich fusions is cotranslational. The release of **A**Ub, **B**Ub, and **C**Ub by UBP cleavage was assayed by immunoprecipitation (*7*). **C**Ub* denotes both the Me^{K}DHFRhaUb band and a cross-reacting band present in these NEM-treated extracts but not in untreated ones (compare with Fig. 3A). M_r, relative molecular mass.

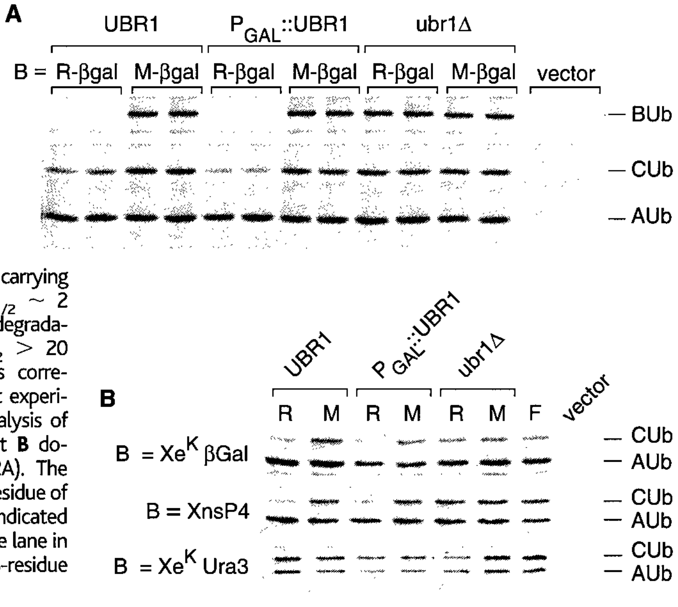


Fig. 3. Nascent polypeptides bearing an N-terminal degradation signal can be degraded cotranslationally. (**A**) Determination of **C**/**A** ratios through immunoprecipitation of in vivo–labeled **A**Ub-**B**Ub-**C**Ub fusion proteins (*14*). Two variants of βgal were used as domain **B**, one carrying an N-degron (Re^K-βgal; $t_{1/2} \sim 2$ min) and one lacking this degradation signal (Me^K-βgal; $t_{1/2} > 20$ hours). Each pair of lanes corresponds to two independent experiments. (**B**) Immunoblot analysis of the **C**/**A** levels for different **B** domains (described in Fig. 2A). The identity of the N-terminal residue of each of the **B** domains is indicated above each lane. F marks the lane in which domain **B** was the 18-residue FLAG-containing moiety.

To verify that UBP-mediated cleavage is cotranslational (*3*), we carried out in vivo radiolabeling in which the labeling pulse was substantially shorter than the time required for the complete synthesis of AUb-**B**Ub-CUb. *Saccharomyces cerevisiae* cells expressing the fusion protein {DHFRhaUb} - {Me^{K}βgalUb} - {Me^{K}DHFRhaUb} (Fig. 2A), predicted to require ~350 s for complete synthesis, were radiolabeled for 45 s (*7*). Labeling was terminated by addition of cycloheximide, and UBPs were simultaneously inactivated with *N*-ethylmaleimide (NEM) (*3*). Under these conditions, nascent chains that are just starting to be synthesized when the pulse begins will incorporate label into the N-terminal **A** domain but do not elongate to full-length chains (Fig. 2B). Thus, detection of free, labeled AUb and **B**Ub by immunoprecipitation indicated that UBP-mediated cleavage at the AUb-**B**Ub junction was cotranslational (Fig. 2C). No full-length AUb-**B**Ub-CUb fusion was detected (Fig. 2C), indicating that cotranslational cleavage by UBPs was highly efficient.

Previous experiments bearing on cotranslational degradation used inhibitors or cell-free systems (*8–10*). Nascent polypeptide chains might be protected from degradation in vivo, either because they are sterically

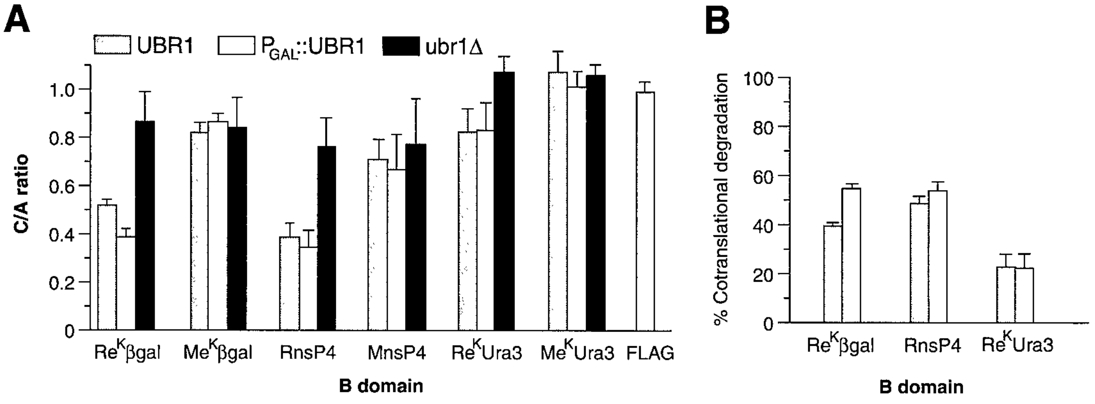


Fig. 4. The extent of cotranslational protein degradation depends on the presence of a degron, the activity of a degron-specific proteolytic pathway, and the nature and size of the protein. (**A**) C/A ratios obtained for the different **B** domains. (**B**) Ubr1p-dependent cotranslational degradation (*15*). Each bar represents a mean value derived from at least four independent experiments; standard errors are indicated. Differences between means were significant to at least $P < 0.05$ by the Mann-Whitney test. Bars are as in (A).

shielded by chaperones or because their translation time is short compared with the time required for targeting by the degradation machinery. The Ub sandwich technique was used to detect in vivo cotranslational degradation of a 118-kD, β-galactosidase (βgal)–derived polypeptide carrying a strong N-terminal degradation signal, specifically an N-degron (Fig. 2A). The βgal-linked N-degron comprises the destabilizing N-terminal residue arginine (R) and a short, lysine (K)–bearing extension, e^K (Fig. 2A), that is the site of multi-Ub chain attachment (*11*). Ubr1p, the E3 component of the N-end rule pathway, targets Re^K-βgal for rapid degradation in vivo [half-life ($t_{1/2}$) ~ 2 min] (*2*, *12*). Changing the N-terminal residue of the protein to methionine (M) inactivates the degradation signal by precluding recognition by Ubr1p. The resulting Me^K–βgal is posttranslationally stable ($t_{1/2}$ > 20 hours) (*2*). **A**Ub-**B**Ub-**C**Ub fusion proteins in which domain **B** was either Re^K-βgal or its N- degron–lacking counterpart Me^K-βgal (Fig. 2A) were expressed in *S. cerevisiae* strains containing different levels of Ubr1p (*13*). The extent of cotranslational degradation was assessed by radiolabeling for 30 min and immunoprecipitation (*14*) to determine the levels of **C**Ub relative to **A**Ub (the C/A ratio).

The C/A ratio was lower in cells expressing the N-degron–bearing domain **B**, but only in those strains that also expressed Ubr1p (Fig. 3A). To determine the percentage of nascent chains cotranslationally degraded by the N-end rule pathway, we compared the C/A ratios for Re^K-βgal in wild-type (0.52) and Ubr1p-overexpressing strains (0.39) with the ratio found with the *ubr1Δ* strain (0.86) (Fig. 4A). This comparison indicated that ~40% of the nascent Re^K-βgal chains were cotranslationally degraded in the wild-type strain (Fig. 4B) (*15*). This fraction increased to ~55% when Ubr1p was overexpressed from the P_{GAL1} promoter.

The extent of cotranslational degradation of two other **B** domains of different sizes was also determined. The mammalian Sindbis virus RNA polymerase, termed nsP4, is a 69-kD protein that naturally bears an N-degron (*16*), and Xe^K-Ura3p (X = M or R) is a 34-kD enzyme of the *S. cerevisiae* uracil biosynthetic pathway that either carries (Re^K-Ura3p) or lacks (Me^K-Ura3p) an N-degron (*17*) (Fig. 2A). Although the short-lived R-nsP4 (69 kD) and Re^K-βgal (118 kD) were cotranslationally degraded to similar extents, Re^K-Ura3p (34 kD), which was also short-lived posttranslationally (*17*), exhibited much less cotranslational degradation than the other two proteins (Fig. 3B). Radiolabeling and immunoprecipitation (*14*) showed that ~50% of the nascent chains of R-nsP4 were cotranslationally degraded by the N-end rule pathway in wild-type cells and ~55% in cells overexpressing Ubr1p (Fig. 4B). These values were similar to ~40% and ~55% cotranslational degradation of Re^K-βgal in these strains, respectively (Fig. 4B). In contrast, only ~20% cotranslational degradation was observed with the 34-kD Re^K-Ura3p in either wild-type or Ubr1p-overexpressing strains (Fig. 4B). These results suggest that smaller proteins are less susceptible to cotranslational degradation; however, factors other than translation time likely influence the presentation or accessibility of the degradation signal by the nascent chain, because protein size was not directly proportional to the extent of cotranslational degradation.

These comparisons established the amounts of cotranslational degradation by the N-end rule pathway. The C/A ratios for two of the N-degron–lacking **B** domains, Me^K-βgal (0.84) and M-nsP4 (0.72), were less than the C/A ratio obtained with a very short FLAG epitope–containing sequence that lacks a degradation signal and exhibits a C/A ratio of 0.97, indistinguishable from the theoretical value of 1.0 (Fig. 4A). This suggests that ~15% of Me^K-βgal and ~25% of M-nsP4 nascent chains might be cotranslationally degraded by a Ubr1p-independent pathway. Premature termination of translation and/or transcription may also contribute to the drop in the C/A ratios for these **B** domains. However, it is unlikely that premature termination of translation accounts entirely for the difference between the C/A ratios for FLAG and M-nsP4, because the codon adaptation index of the M-nsP4 open reading frame (0.1) is higher than the one for Me^K-βgal (0.07), but the drop in C/A ratio is greater with M-nsP4 than with Me^K-βgal. The degradation of newly synthesized polypeptides is a major source of peptides presented to the immune system (*18*, *19*). Our observations with the above two **B** domains suggest that the source of these peptides is cotranslational protein degradation. In this regard, it is interesting that nsP4 is a viral protein that is presented efficiently to the immune system by infected cells.

The extent of cotranslational degradation can be strikingly high: In the case of Re^K-βgal in Ubr1p-overexpressing cells, over 50% of nascent polypeptide chains never reach their full size before their destruction by processive proteolysis. Thus, if a nascent chain displays a degron of the Ub system, such a protein becomes a target of kinetic competition between cotranslational biogenesis and cotranslational degradation. Because the folding of a protein molecule begins during its synthesis on the ribosome, a nascent polypeptide may cotranslationally expose degradation signals that become shielded through the folding of the newly formed protein (*20*, *21*). Thus, cotranslational protein degradation may represent a form of protein quality control that destroys nascent chains that fail to fold correctly rapidly enough.

References and Notes

1. A. Varshavsky, *Trends Biochem. Sci.* **22**, 383 (1997).
2. A. Bachmair, D. Finley, A. Varshavsky, *Science* **234**, 179 (1986).
3. N. Johnsson and A. Varshavsky, *EMBO J.* **13**, 2686 (1994).
4. F. Lévy, N. Johnsson, T. Rumenapf, A. Varshavsky, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4907 (1996).
5. The Ub moieties of the fusion protein serve as cleavage sites between domains **A**, **B**, and **C**. They do not target the attached protein for degradation (*4*), nor do they contribute to multi-Ub chain formation, because they carry a lysine to arginine substitution at position 48 (Ub^{R48}) that prevents their conjugation to other Ub molecules at that position (*11*).
6. W. Baumeister, J. Walz, F. Zühl, E. Seemüller, *Cell* **92**, 367 (1998).
7. The expression of Ub sandwich proteins from the

P_{GAL1} promoter on p426GAL1 vector (*22*) was induced with 0.1 μM β-estradiol with a Gal4.ER.VP16 chimeric protein (*23*). Ten-milliliter cell cultures were harvested after 10 hours of induction (absorbance at 600 nm = 0.5 to 1) and labeled for 30 s with ^{35}S-EXPRESS (0.7 mCi/ml; New England Nuclear). Cycloheximide and NEM (final concentrations of 0.5 mg/ml and 0.1 M, respectively) were then added, cell extracts were prepared from 0.2 ml of the resulting sample, and immunoprecipitations were carried out as described (*17*), with antibodies to ha (12CA5 Boehringer), βgal (Promega), and/or FLAG (M2; Eastman Kodak), as appropriate. Total handling time to the beginning of cell lysis was 45 s. In the eukaryotic cell types examined, the rate of translation is 2 to 10 residues per second (*24*). Assuming a rate of five residues per second in *S. cerevisiae*, this pulse is seven times shorter than the duration of translation of the fusion protein. Immunoprecipitates were fractionated by 13% SDS–polyacrylamide gel electrophoresis, and band intensities were quantitated by PhosphorImager (Molecular Dynamics).

8. L. Lin, G. N. DeMartino, W. C. Greene, *Cell* **92**, 819 (1998).
9. W. Liao, S. C. J. Yeung, L. Chan, *J. Biol. Chem.* **273**, 27225 (1998).
10. S. Sato, C. L. Ward, R. R. Kopito, *J. Biol. Chem.* **273**, 7189 (1998).
11. V. Chau *et al.*, *Science* **243**, 1576 (1989).
12. A. Varshavsky, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12142 (1996).
13. The *S. cerevisiae* strains JD52 (*MAT***a** *lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-3,112*), JD54 (P_{GAL1}*::UBR1*), and JD55 (*ubr1Δ*) were described (*17*).
14. The procedure was as in (*7*), except that cells were labeled for 30 min with 0.11 mCi of ^{35}S-EXPRESS, then pelleted, resuspended in 0.8 ml of ice-cold lysis buffer, and processed for immunoprecipitation (*17*).
15. For example, with Re^{K}-βgal, where C/A was 0.52 in wild-type cells and 0.86 in *ubr1Δ*, the level of the Ubr1p-dependent cotranslational degradation was 100% × (1 − 0.52/0.86) = 40%.
16. R. J. deGroot, T. Rümenapf, R. J. Kuhn, J. H. Strauss, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8967 (1991).
17. M. Ghislain, R. J. Dohmen, F. Lévy, A. Varshavsky, *EMBO J.* **15**, 4884 (1996).
18. E. A. J. Reits, J. C. Vos, M. Gromme, J. Neefjes, *Nature* **404**, 774 (2000).
19. U. Schuberg *et al.*, *Nature* **404**, 770 (2000).
20. J. Frydman and F. U. Hartl, *Science* **272**, 1497 (1996).
21. B. Bukau, E. Deuerling, C. Pfund, E. A. Craig, *Cell* **101**, 119 (2000).
22. D. Mumberg, R. Muller, M. Funk, *Nucleic Acids Res.* **22**, 5767 (1994).
23. J. F. Louvion, B. Havauxcopf, D. Picard, *Gene* **131**, 129 (1993).
24. A. S. Spirin, *Ribosome Structure and Protein Biosynthesis* (Cummings, Menlo Park, CA, 1986).

25. We thank F. Lévy and N. Johnsson for their contributions to early stages of this work and R. Deshaies, T. Iverson, T.-M. Yi, and members of the Varshavsky laboratory for helpful discussions and comments on the manuscript. This study was supported by a grant to A.V. from NIH. G.T. was supported in part by Amgen.

15 March 2000; accepted 31 July 2000

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 275, No. 30, Issue of July 28, pp. 22931–22941, 2000
Printed in U.S.A.
© 2000 by The American Society for Biochemistry and Molecular Biology, Inc.

RGS4 Is Arginylated and Degraded by the N-end Rule Pathway *in Vitro**

Received for publication, February 28, 2000, and in revised form, April 22, 2000
Published, JBC Papers in Press, April 26, 2000, DOI 10.1074/jbc.M001605200

Ilia V. Davydov and Alexander Varshavsky‡
From the Division of Biology, California Institute of Technology, Pasadena, California 91125

The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. We used an expression-cloning screen to search for mouse proteins that are degraded by the ubiquitin/proteasome-dependent N-end rule pathway in a reticulocyte lysate. One substrate thus identified was RGS4, a member of the RGS family of GTPase-activating proteins that down-regulate specific G proteins. A determinant of the RGS4 degradation signal (degron) was located at the N terminus of RGS4, because converting cysteine 2 to either glycine, alanine, or valine completely stabilized RGS4. Radiochemical sequencing indicated that the N-terminal methionine of the lysate-produced RGS4 was replaced with arginine. Since N-terminal arginine is a destabilizing residue not encoded by *RGS4* mRNA, we conclude that the degron of RGS4 is generated through the removal of N-terminal methionine and enzymatic arginylation of the resulting N-terminal cysteine. RGS16, another member of the RGS family, was also found to be an N-end rule substrate. RGS4 that was transiently expressed in mouse L cells was short-lived in these cells. However, the targeting of RGS4 for degradation in this *in vivo* setting involved primarily another degron, because N-terminal variants of RGS4 that were stable in reticulocyte lysate remained unstable in L cells.

A multitude of regulatory circuits, including those that control the cell cycle, cell differentiation, and responses to stress, involve metabolically unstable proteins (1–5). A short *in vivo* half-life of a regulator provides a way to generate its spatial gradients and allows for rapid adjustments of its concentration, or subunit composition, through changes in the rate of its synthesis or degradation. Damaged or otherwise abnormal proteins tend to be short-lived as well (6). Features of proteins that confer metabolic instability are called degradation signals or degrons (7, 8). The essential component of one degradation signal, called the N-degron, is a destabilizing N-terminal residue of a protein (9). A set of amino acid residues that are destabilizing in a given cell yields a rule, called the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. The N-end rule pathway is present in all organisms examined, from mammals and plants to fungi and prokaryotes (10, 11).

In eukaryotes, an N-degron consists of two determinants, a destabilizing N-terminal residue and an internal lysine or lysines (12–15). The Lys residue is the site of formation of a multiubiquitin chain (16). The N-end rule pathway is thus one of the pathways of the ubiquitin (Ub)[1] system. Ub is a 76-residue protein whose covalent conjugation to other proteins plays a role in a vast range of biological processes (4, 5, 17). In most of them, Ub acts through routes that involve the degradation of ubiquitylated[2] proteins by the 26 S proteasome, an ATP-dependent multisubunit protease (18, 19).

The N-end rule has a hierarchic structure. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their deamidation, by the *NTA1*-encoded[3] N-terminal amidase (Nt-amidase), to yield the secondary destabilizing residues Asp and Glu (20). The destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *ATE1*-encoded Arg-tRNA-protein transferase (R-transferase), to Arg, one of the primary destabilizing residues (21, 22). The primary destabilizing N-terminal residues are bound directly by UBR1, also called N-recognin, the E3 (recognition) component of the N-end rule pathway (10, 11).

In mammals, the deamidation step is mediated by two Nt-amidases, Nt^N-amidase and Nt^Q-amidase, which are specific, respectively, for N-terminal Asn and Gln (23, 24). In vertebrates, the set of secondary destabilizing residues contains not only Asp and Glu but also Cys, which is a stabilizing residue in yeast (25, 26). The mammalian counterpart of the yeast R-transferase Ate1p exists as two distinct species, ATE1–1 and ATE1–2, that are produced through alternative splicing of *Ate1* pre-mRNA (27). Both ATE1–1 and ATE1–2 are similar in specificity to the *ATE1*-encoded yeast R-transferase, in that these R-transferases can arginylate N-terminal Asp and Glu, but cannot arginylate N-terminal Cys (27), suggesting the existence of a distinct R-transferase specific for N-terminal Cys.

---

* This work was supported by Grants GM31530 and DK39520 (to A. V.) from the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed: Division of Biology, 147–75, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125. Tel.: 626-395-3785; Fax: 626-440-9821; E-mail: avarsh@caltech.edu.

[1] The abbreviations used are: Ub, ubiquitin; Nt-amidase, N-terminal amidase; R-transferase, Arg-tRNA-protein transferase; Ub-*X*-nsP4*, Ub-*X*-nsP4$_{1-254}$; GST, glutathione transferase; RGS, regulator of G-protein signaling; MetAP, methionine aminopeptidase; GAP, GTPase-activating protein; DUB, deubiquitylating enzyme; AMP-PNP, 5′-adenylylimidodiphosphate; CAPS, 3-(cyclohexylamino)propanesulfonic acid; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; ORF, open reading frame.

[2] Ubiquitin whose C-terminal (Gly-76) carboxyl group is covalently linked to another compound is called the *ubiquityl* moiety, the derivative terms being *ubiquitylation* and *ubiquitylated*. The term Ub refers to both free ubiquitin and the ubiquityl moiety. This nomenclature (5), which is also recommended by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (74), brings ubiquitin-related terms in line with the standard chemical terminology.

[3] Throughout the text, the names of genes are italicized and all uppercase. The names of proteins are roman and all uppercase. This usage, a modification of the existing conventions (73), provides a uniform nomenclature in a text that refers to both fungal and metazoan genes and proteins.

UBR1 (N-recognin) of yeast and mammals has two binding sites for the primary destabilizing N-terminal residues of either proteins or short peptides. The type 1 site is specific for the basic N-terminal residues Arg, Lys, and His. The type 2 site is specific for the bulky hydrophobic N-terminal residues Phe, Leu, Trp, Tyr, and Ile (25, 28, 29). UBR1 contains yet another substrate-binding site, which targets proteins bearing internal (non-N-terminal) degrons. These proteins include CUP9 and GPA1 in yeast (30–32) and the encephalomyocarditis virus 3C protease in metazoans (33).

The known functions of the N-end rule pathway include the control of peptide import in *S. cerevisiae*, through the degradation of CUP9, a transcriptional repressor of the peptide transporter PTR2 (30) (this control includes a positive feedback mediated by the type 1 and type 2 sites of UBR1 (31)); the degradation of GPA1, one of two Gα proteins in *S. cerevisiae* (32); and the degradation of alphaviral RNA polymerases and other viral proteins in infected metazoan cells (33, 34). Physiological N-end rule substrates were also identified among the proteins secreted into the cytosol of the mammalian cell by intracellular parasites such as the bacterium *Listeria monocytogenes* (35). Selective perturbation of the N-end rule pathway was reported to interfere with mammalian cell differentiation (36, 37) and with limb regeneration in amphibians (38). Studies of the Ub-dependent proteolysis of endogenous proteins in muscle extracts suggested that the N-end rule pathway plays a role in catabolic states that result in muscle atrophy (39).

Until the present work, physiological substrates of Nt-amidases and R-transferases were unknown in either yeast or larger eukaryotes. Engineered N-end rule substrates, including the substrates of Nt-amidases and R-transferases, can be produced *in vivo* through the Ub fusion technique, in which a Ub-X-reporter fusion is cleaved, cotranslationally, after the last residue of Ub by deubiquitylating enzymes (DUBs) (40), yielding a reporter protein bearing the predetermined N-terminal residue X (9, 10, 41).

In the present work, we employed a modification of the cDNA-based sib-selection strategy in a transcription-translation lysate from rabbit reticulocytes (42, 43) to identify putative physiological substrates of the N-end rule pathway. Specifically, we used dipeptides bearing destabilizing N-terminal residues as selective inhibitors of the N-end rule pathway, and we screened for mouse cDNAs that expressed proteins whose relative abundance in the lysate was altered in the presence of relevant dipeptides.

Among the putative N-end rule substrates identified through the use of this approach was mouse RGS4, a GTPase-activating protein (GAP) for specific Gα subunits of heterotrimeric G proteins, and a member of the family of RGS (regulator of G protein signaling) proteins (44–50). We discovered that in addition to the expected removal of N-terminal Met from the newly formed RGS4, the resulting N-terminal Cys of RGS4 was arginylated, presumably by a distinct R-transferase whose Cys specificity is different from that of the known R-transferases. Thus modified RGS4 bore N-terminal Arg, a primary destabilizing residue, and was degraded by the N-end rule pathway in reticulocyte lysate.

EXPERIMENTAL PROCEDURES

Plasmids—The plasmids pcDNA3-Ub-X-nsP4$_{1-254}$ (where X is Met, Arg, or Tyr) were used to express Ub-X-nsP4$_{1-254}$ fusions, denoted below as Ub-X-nsP4*, from the phage T7 promoter in the transcription-translation reticulocyte lysate system by Promega (Madison, WI). The notation nsP4$_{1-254}$ refers to the 254-residue N-terminal fragment of the 69-kDa nsP4, the Sindbis virus RNA polymerase (34). The pcDNA3-Ub-X-nsP4$_{1-254}$ plasmids were constructed using a set of open reading frames (ORFs) encoding fusions between Ub and the full-length nsP4 (51). The Ub-X-nsP4 ORFs in pJCEX1 (a gift from Dr. T. Rümenapf, Federal Research Center for Virus Diseases of Animals, Tübingen, Germany) were used as templates for PCR with two primers 5′-TTCGGATCCGCCACCATGCAGATCTTCGTGAAGACCCTG-3′ and 5′-CCTTCTAGACTATGCGGTGACAAACTCAGTGGTAAT-3′, to produce DNA fragments encoding Ub-X-nsP4$_{1-254}$ fusions (the underlined sequences corresponded to the start and stop codons, respectively). These fragments were digested with *Bam*HI and *Xba*I and cloned into the *Bam*HI/*Xba*I-cut pcDNA3 vector (Invitrogen, Carlsbad, CA).

The plasmid pcDNA3-Ub-Lys-mCL expressed Ub-Lys-mCL, where Lys-mCL was an N-terminally truncated large subunit of m-calpain starting with Lys-10. The primers 5′-ATTCCGCGGTGGCAAAGACCGCGAGGCGGCCGAGGGGCTG-3′ and 5′-CCTACTAGTCTATCATAGGACTGAAAAACTCAGCCACGAGAT-3′ were used to amplify a DNA fragment from pET24–80k, which contained cDNA encoding the large subunit of rat m-calpain (52) (a gift from Dr. J. S. Elce, Queen's University, Kingston, Ontario, Canada). The resulting fragment was digested with *Sac*II and *Spe*I and cloned into *Sac*II/*Spe*I-cut pcDNA3-Ub-Met-nsP4$_{1-254}$. The plasmids pcDNA3-p94 and pcDNA3-p94$_{C129A}$ were used to express the mouse-specific calpain p94 and its C129A mutant. To produce these constructs, plasmids containing the rat p94 ORF and the p94$_{C129A}$ ORF (53) (gifts from Dr. H. Sorimachi and Dr. K. Suzuki, University of Tokyo, Japan) were PCR-amplified and cloned into the pcDNA3 vector.

The plasmid pcDNA-RGS4 was used for expression of the mouse RGS4 in reticulocyte lysate and for transfection-mediated expression of RGS4 in mouse L cells. This plasmid contained the T7 and CMV promoters, and the mouse RGS4 ORF followed by two stop codons. The RGS4 ORF was PCR-amplified from the plasmid pcDNA3-26-16-11 (see below), using the primers 5′-TTCGGATCCGCCACCATGTGCAAAGGACTTGCAGGTCTG-3′ and 5′-CCTTCTAGATCATTAGGCACACTGGGAGACCAGGGA-3′, followed by digestion with *Bam*HI and *Xba*I and cloning into *Bam*HI/*Xba*I-cut pcDNA3 (the underlined sequences corresponded to the start codon and two stop codons, respectively). The plasmids pcDNA3-RGS4$_{M19A}$, pcDNA3-RGS4$_{C2G}$, pcDNA3-RGS4$_{C2V}$, pcDNA3$_{C2A}$, pcDNA3-RGS4$_{K3R}$, pcDNA3-RGS4$_{K3S}$, and pcDNA3-RGS4$_{19-205}$, which expressed specific RGS4 mutants, were constructed from pcDNA3-RGS4 using PCR-based site-directed mutagenesis (54).

The plasmid pcDNA3-RGS4$_{M19A}$-GST expressed RGS4$_{M19A}$-GST, a fusion of RGS4$_{M19A}$ and glutathione transferase (GST). It was constructed through a PCR-mediated fusion of DNA fragments encoding RGS4 and GST, followed by cloning into *Bam*HI/*Xba*I-cut pcDNA3. PCR-mediated site-directed mutagenesis was then used to introduce the M19A mutation. The plasmid pcDNA3-RGS16, expressing mouse RGS16 from the T7 promoter (55), and a PCR-produced fragment encoding mouse Gβ_{5L} were gifts from Dr. C. K. Chen and Dr. M. I. Simon (California Institute of Technology, Pasadena, CA). The entire coding regions of the final plasmid constructs were verified by DNA sequencing.

In Vitro Transcription-Translation-Degradation System—The TNT Quick-coupled Transcription-Translation System (Promega) contained a rabbit reticulocyte lysate pre-mixed with most of the reaction components necessary to carry out transcription/translation in the lysate, including all of the amino acids except methionine. [^{35}S]Methionine (>1,000 Ci/mmol, Amersham Pharmacia Biotech) was used to label newly formed proteins in the lysate. Proteins labeled with ^{3}H-amino-acids were produced in the TNT T7-coupled Reticulocyte Lysate (Promega), a version of the system where the main components of the reaction were supplied separately. The reactions were set up according to the manufacturer's instructions. The reaction mixtures containing dipeptides also contained 0.15 mM bestatin (Sigma), an inhibitor of some aminopeptidases, to reduce degradation of the added dipeptides (28). Stock samples of dipeptides were 0.5 M solutions in 10 mM K-HEPES, pH 7.5. The reactions were incubated at 30 °C, unless stated otherwise; they were terminated by the addition of SDS-sample buffer, heated at 95 °C for 5 min, and fractionated by SDS-PAGE, followed by autoradiography or fluorography. ^{35}S in protein bands was determined using PhosphorImager (Molecular Dynamics, Sunnyvale, CA).

Screening of Mouse cDNA Library of Small Pools in Rabbit Reticulocyte Lysate—Total RNA was isolated from the brains of female mice with the guanidine thiocyanate method (54). Poly(A)$^{+}$ RNA was purified using Oligotex mRNA kit (Qiagen, Valencia, CA) and was used to synthesize oligo(dT)-primed double-stranded cDNA with the SuperScript cDNA synthesis kit (Life Technologies, Inc.). After ligating the DNA fragments to a *Bst*XI adapter (Invitrogen, San Diego) and digesting with *Not*I, the cDNAs were cloned into *Bst*XI-*Not*I-cut pcDNA3. The resulting cDNA library was introduced, by electroporation, into *Escherichia coli* SURE cells (Stratagene, La Jolla, CA), and the transformants were frozen in a number of samples. Titration of the library showed that it contained 3–4 × 10^6 independent clones. Pools of the library cDNAs containing each about 50 clones were prepared as described (42). Individual pools were added to the T7 TNT Quick-coupled Transcription-Translation System (Promega) together with [^{35}S]methionine, in either the presence or the absence of the mixture of 5 mM Arg-β-Ala, 5 mM Trp-Ala and 0.15 mM bestatin. All components of a reaction except the lysate were gently mixed together, followed by the addition of lysate. The reactions were performed in the total volume of 6.25 μl in a 96-well microtiter plate at 30 °C for 3.5 h. These conditions were chosen after preliminary optimization, using Tyr-nsP4* as a test protein. Once a cDNA pool was found to express a protein whose relative abundance was increased in the presence of the dipeptides, the pool was progressively subdivided and retested, until the isolation of a single positive cDNA clone. One of the clones thus isolated, termed pcDNA3-26-16-11, contained as ~3-kilobase pair mouse RGS4 cDNA (see "Results").

N-terminal Radiochemical Sequencing—RGS4$_{M19A}$-GST was expressed in the TNT T7-coupled Reticulocyte Lysate System (Promega) in the presence of either [^{3}H]arginine (49 Ci/mmol), [^{3}H]leucine (155 Ci/mmol), or [^{3}H]lysine (87 Ci/mmol) (Amersham Pharmacia Biotech), 1 mM Arg-β-Ala, and 0.15 mM bestatin for 30 min at 30 °C, in the total volume of 0.4 ml. The reaction was stopped by the addition of 20 mM AMP-PNP (Sigma), a non-hydrolyzable ATP analog, and the labeled RGS4$_{M19A}$-GST was partially purified by affinity chromatography on glutathione-Sepharose 4B (Amersham Pharmacia Biotech), followed by SDS-10% PAGE and the electrophoretic transfer of separated proteins in 10% methanol, 10 mM CAPS, pH 11, to Immobilon-P membrane (Millipore, Burlington, MA). The transferred band of RGS4$_{M19A}$-GST was cut out (its position was determined with respect to stained protein markers) and subjected to multiple cycles of Edman degradation, using the 476A Applied Biosystems sequencer (Perkin-Elmer). The fractions from each cycle were collected, and ^{3}H in the fractions was determined using a scintillation counter.

Transfections of Mouse L Cells and Pulse-Chase Assay—Transient transfections of mouse L cells and pulse-chase analysis were performed as described previously (56). The affinity-purified polyclonal antibody raised against a peptide near the C terminus of RGS4 was purchased from Santa Cruz Biotechnology (Santa Cruz, CA). In the experiments with proteasome inhibitor MG132 (Calbiochem), it was added to cells, at the final concentration of 50 μM, 45 min before the addition of [^{35}S]methionine and was present throughout the pulse-chase assay.

RESULTS

Degradation of N-end Rule Substrates in a Transcription-Translation System Derived from Reticulocytes—The N-end rule pathway can be selectively inhibited in reticulocyte lysates through the addition of dipeptides bearing either type 1 or type 2 destabilizing N-terminal residues (25, 28). A commercially produced transcription-translation system derived from reticulocyte lysate and containing the phage T7 RNA polymerase was used to express a protein of interest from DNA template in one reaction mixture. A putative N-end rule substrate in the lysate was detected through its increased concentration in the presence of a cognate dipeptide inhibitor of the N-end rule pathway. We began by examining the degradation of model N-end rule substrates, which were derived from nsP4, the 69-kDa Sindbis virus RNA polymerase, a physiological substrate of the mammalian N-end rule pathway that bears N-terminal Tyr (34). The nsP4 protein is produced during Sindbis virus infection through site-specific cleavage of the viral polyprotein precursor. In the present work, the Ub fusion technique (9, 10, 51) was used to synthesize, in reticulocyte lysate, a set of *X*-nsP4 derivatives that contained the first 254 residues of nsP4 and in addition bore different residues at the Ub-nsP4 junction. These test proteins, Ub-Arg-nsP4$_{1-254}$, Ub-Tyr-nsP4$_{1-254}$, and Ub-Met-nsP4$_{1-254}$, are denoted below as Ub-Arg-nsP4*, Ub-Tyr-nsP4*, and Ub-Met-nsP4*, respectively. Their N-terminal Ub moiety was cotranslationally cleaved off by DUBs in the lysate yielding Tyr-nsP4*, Arg-nsP4*, and Met-nsP4*, respectively. Earlier experiments by T. Rümenapf have shown that these constructs were N-end rule substrates in the transcription-translation reticulocyte lysate.[4]

When a Ub-*X*-nsP4* fusion was expressed in reticulocyte lysate and monitored as a function of time, a major band corresponding to Ub-lacking *X*-nsP4* was observed (Fig. 1, *A* and *C*). Although the electrophoretic mobility of *X*-nsP4* proteins was slightly higher than expected from their predicted molecular mass of 29 kDa, the observed proteins were clearly *X*-nsP4*, rather than, for example, Ub-*X*-nsP4*, because the removal of Ub occurs cotranslationally or nearly so (57), and also because SDS-PAGE of the same samples using a more concentrated polyacrylamide gel revealed the ~8-kDa band of labeled free Ub (data not shown). In addition, the degradation patterns of presumed *X*-nsP4* derivatives of Ub-*X*-nsP4* fusions conformed to the N-end rule, as shown below.

The metabolically stable Met-nsP4* accumulated rapidly during the first 30 min of the transcription-translation reaction and reached a plateau around 30 min, because the protein synthesis (but not the activity of the N-end rule pathway; see below) ceased in the lysate by that time (Fig. 1, *C* and *D*). Thus, in this setting, the incubation times in excess of ~30 min were operationally equivalent to the "chase" part of a pulse-chase experiment. In contrast to Met-nsP4*, the relative levels of Arg-nsP4* (Fig. 1, *A* and *B*, *lanes* labeled *c* (controls)) and Tyr-nsP4* (data not shown) began to decrease after 30 min, reflecting their continuing degradation by the N-end rule pathway. Arg and Tyr are, respectively, a type 1 (basic) and a type 2 (bulky hydrophobic) destabilizing residue in the N-end rule (10). A ladder of higher molecular mass bands, apparently of multiubiquitylated Arg-nsP4*, was observed with Arg-nsP4* (Fig. 1*A*) but not with the metabolically stable Met-nsP4* (data not shown).

The addition of Arg-β-Ala, a type 1 inhibitor of the N-end rule pathway (25), strongly inhibited the ubiquitylation and degradation of Arg-nsP4* (Fig. 1*A*, *lanes* labeled *1*). The same dipeptide had no effect on the degradation of Tyr-nsP4*, which bore a type 2 destabilizing N-terminal residue (data not shown). Conversely, Trp-Ala, a type 2 inhibitor, strongly decreased the degradation of Tyr-nsP4* (Fig. 2*B*) but had little effect on the degradation of Arg-nsP4* (Fig. 1*A*, *lanes* labeled *2*). At 1 mM, none of the several dipeptides tested, including those bearing type 1, type 2, or type 3 destabilizing N-terminal residues (25), affected the relative band intensity of Met-nsP4* (Fig. 1, *C* and *D*), indicating that these dipeptides did not perturb transcription and translation in this system. However, at 10 mM, some of the dipeptides significantly delayed the synthesis of *X*-nsP4* (data not shown). Therefore the experiments were carried out with dipeptides at the initial concentration of 1 mM. At this concentration, the inhibition of the N-end rule pathway was strong but incomplete. The inhibition of degradation of, for example, a type 1 substrate such as Arg-nsP4* by a cognate dipeptide could be observed through large differences in the amount of a test protein produced by 30 min, the time of cessation of protein synthesis in the lysate, and through even larger differences at 60 min (Fig. 1*B*).

m-Calpain Bearing N-terminal Lysine Is Not a Substrate of the N-End Rule Pathway—m-Calpain is a calcium-activated cysteine protease composed of an 80-kDa large subunit (mCL) and a 30-kDa small subunit (mCS) (58). The m- and μ-calpains are ubiquitously expressed in metazoan cells. In the presence of Ca^{2+}, calpain subunits undergo autoproteolytic cleavages in their N-terminal regions. The processed calpains have increased activity. The *in vitro* autoproteolytic cleavage of mCL was shown to remove the first 9 residues of mCL, yielding a

---

[4] T. Rümenapf and A. Varshavsky, unpublished data.

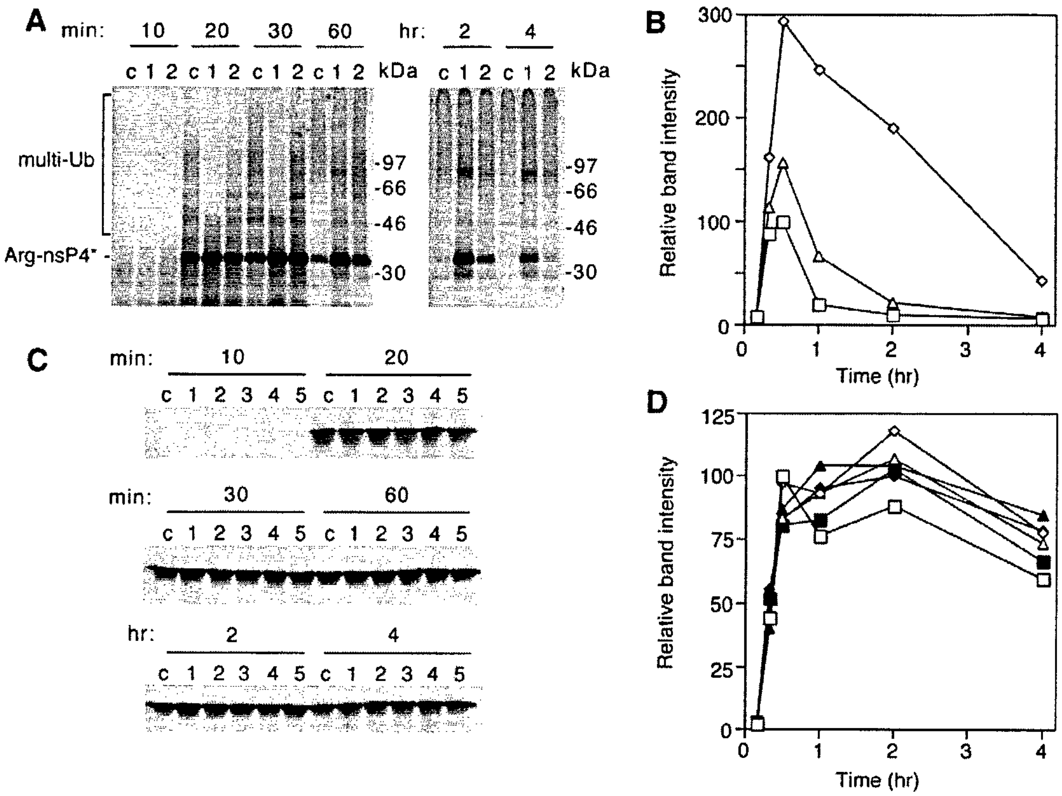


FIG. 1. **Effects of dipeptides on the accumulation of Arg-nsP4* and Met-nsP4* in the transcription-translation reticulocyte lysate.** Plasmids expressing Ub-Arg-nsP4* (*A*) or Ub-Met-nsP4* (*C*) were added to reticulocyte lysate in the presence of [^{35}S]methionine and either in the absence of N-end rule inhibitors (*lane c* (controls)) or in the presence of either 1 mM Arg-β-Ala (*lanes 1*), or 1 mM Trp-Ala (*lanes 2*), or 1 mM Ala-Lys (*lanes 3*), or 1 mM Lys-Ala (*lanes 4*), or 1 mM Ala-Arg (*lanes 5*). The samples in *lanes 1–5* also contained 0.15 mM bestatin. The reactions were carried out for 10, 20, 30, 60, 120, and 240 min, followed by SDS-PAGE, autoradiography, and quantitation. The position of R-nsP4* band and the multi-Ub-containing species are indicated on the *left*. The positions of molecular mass markers are indicated on the *right*. *B*, relative amounts of ^{35}S-labeled Arg-nsP4* in *A* were determined using PhosphorImager and were plotted as a function of incubation time. The amount of Arg-nsP4* in the reaction without dipeptides at 30 min was assigned the value of 100. Control reaction are as follows: in the absence of dipeptides (□); in the presence of Arg-β-Ala (◇); in the presence of Trp-Ala (Δ); in the presence of Ala-Lys (■); in the presence of Lys-Ala (♦); and in the presence of Ala-Arg (▲). *D*, relative amounts of [^{35}S]Met-nsP4* in *C* were determined and plotted as in *B*.

modified subunit, denoted below as Lys-mCL, which bears N-terminal Lys (59), a type 1 destabilizing residue (10). We asked whether Lys-mCL was unstable in reticulocyte lysate and, if so, whether its instability required N-terminal Lys residue. A plasmid encoding the Ub-Lys-mCL fusion was constructed, and the fusion was expressed in the transcription-translation lysate. It was found that under standard conditions (in the absence of added Ca^{2+}), the fusion-derived Lys-mCL was metabolically stable in the lysate (Fig. 2*A*). In the presence of added Ca^{2+} (2 mM), the newly formed Lys-mCL underwent rapid (apparently autolytic) degradation in the lysate. Significantly, this Ca^{2+}-induced degradation of Lys-mCL was not inhibited by Arg-β-Ala, a type 1 inhibitor of the N-end rule pathway (Fig. 2). Thus, the Lys-mCL derivative of the large subunit of m-calpain is not a substrate of the N-end rule pathway in reticulocyte lysate, despite the presence of a destabilizing N-terminal residue (see "Discussion").

The muscle-specific calpain p94 is homologous to the ubiquitous calpains and was shown to be metabolically unstable (53, 60). We expressed rat p94 and its proteolytically inactive mutant $p94_{C129A}$ in reticulocyte lysate. The full-length p94 was rapidly cleaved in the lysate, yielding several fragments, whereas proteolytically inactive $p94_{C129A}$ remained stable (data not shown), in agreement with the earlier evidence (53). Dipeptides Arg-β-Ala and Trp-Ala had no effect on the relative yields of bands corresponding to either the full-length p94 or the products of its autolysis (data not shown). Thus, neither p94 nor its autolytically produced fragments are the substrates of the N-end rule pathway.

Degradation of RGS4 in Reticulocyte Lysate Is Decreased by Type 1 Inhibitors of the N-end Rule Pathway and by a Proteasome Inhibitor—To search for putative N-end rule substrates in a systematic way, we employed a modification of the previously developed *in vitro* screening method that involves the use of small pools of cDNA clones (42, 43). A mouse brain cDNA library was constructed, and about 500 mouse cDNA pools, each containing about 50 cDNAs, were tested, sequentially, in the transcription-translation reticulocyte lysate. Each reaction was carried out in either the absence or the presence of a mixture of two dipeptides, Arg-β-Ala and Trp-Ala, which have been shown to inhibit selectively the degradation of engineered N-end rule substrates in the lysate (Figs. 1, *A* and *B*, and 2*B*). The resulting ^{35}S-labeled proteins were visualized by SDS-PAGE and autoradiography. We searched for protein bands that were selectively enhanced in the presence, but not in the absence, of the dipeptide inhibitors. A cDNA pool containing a putative substrate of the N-end rule pathway was subjected to subcloning, followed by re-analysis in the lysate. This procedure yielded specific cDNAs encoding putative N-end rule substrates.

Thus far, we found 7 cDNAs that encoded proteins whose expression patterns identified them as putative substrates of the N-end rule pathway. Most of these cDNAs encoded N-terminally truncated polypeptides, produced by translation from Met codons that were internal in the corresponding wild type ORFs (data not shown). Although some of these truncated proteins may prove to be physiologically relevant substrates of the N-end rule pathway, we focused at first on a putative N-end

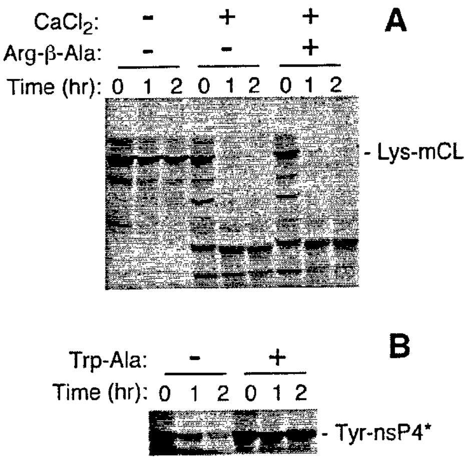


FIG. 2. **Lys-mCL is not an N-end rule substrate in reticulocyte lysate.** *A*, plasmid DNA encoding a Ub-Lys-mCL, a Ub fusion to a derivative of the large m-calpain subunit bearing N-terminal Lys (see "Results"), was expressed in the transcription-translation reticulocyte lysate in the presence of [^{35}S]methionine. The reaction was allowed to proceed for 20 min at 37 °C ("pulse"). Thereafter, cycloheximide and unlabeled L-methionine were added to all samples, to the final concentrations of 0.1 mg/ml and 0.5 mM, respectively. $CaCl_2$ and/or Arg-β-Ala were added to some of the samples, to the final concentrations of 2 and 2.5 mM, respectively. 0.15 mM bestatin was present in these samples as well. The reactions mixtures were incubated for the indicated chase times of 0, 1, or 2 h, followed by SDS-PAGE and autoradiography. *B*, plasmid DNA encoding Ub-Tyr-nsP4* was expressed in the lysate, and the reactions were allowed to proceed as in *A* except that Trp-Ala was added, instead of Arg-β-Ala, after the pulse. The positions of (deubiquitylated) Lys-mCL and Tyr-nsP4* bands are as indicated at the *right*.

rule substrate encoded by a full-length cDNA, which was found to encode the mouse RGS4, a member of the RGS family of GAPs for specific Gα subunits of G proteins (44, 45). The ~3-kilobase pair *RGS4* mRNA contained the ~2-kilobase pair long 3′-untranslated region. To optimize the expression of RGS4 in reticulocyte lysate, we constructed the plasmid pcDNA3-RGS4, which contained the full-length *RGS4* ORF, followed by two stop codons, and lacked the wild type 3′-untranslated region of *RGS4* mRNA. As with the original *RGS4* cDNA isolate (data not shown), the transcription-translation of pcDNA3-RGS4 yielded two protein bands, of 25 and 22 kDa (Fig. 3, *A* and *D*). The larger protein was the full-length RGS4, whereas the smaller one was produced by translation that started from a downstream Met (AUG) codon at amino acid position 19, because pcDNA3-$RGS4_{M19A}$, in which the Met-19 codon was converted to an Ala codon, did not express the lower protein band (Fig. 3*D*).

The RGS4 cDNA was initially picked by this screen because upon the incubation times of 2–4 h, the lower (22-kDa) band increased in intensity in the presence, but not in the absence, of dipeptide inhibitors. However, this effect was later found to be unrelated to the N-end rule pathway, as it was observed in the presence of all dipeptides tested, including those that did not inhibit the pathway (Fig. 3*A*). In contrast, at the incubation times from 20 min to 1 h, the relative amount of the full-length RGS4 increased up to 5-fold in the presence of either Arg-β-Ala or Lys-Ala, the type 1 inhibitors of the N-end rule pathway (Fig. 3, *A* and *C*), but not in the presence of Trp-Ala (type 2 inhibitor of the same pathway), or Ala-Lys and Ala-Arg (type 3 inhibitors) RGS4 (Fig. 4). A ladder of high molecular mass bands, apparently of multiubiquitylated RGS4, was observed above the RGS4 and $RGS4_{M19A}$ bands at 10, 20, and 30 min of incubation (Fig. 3, *A* and *B* and data not shown). Arg-β-Ala and Lys-Ala, but not the other tested dipeptides, delayed the appearance of multiubiquitylated RGS4 (Fig. 3, *A* and *B*). The temporal pattern of inhibition of degradation of RGS4 by dipeptide inhibitors was similar to that described (and explained) above for the engineered N-end rule substrate Arg-nsP4* (Fig. 1, *A* and *B*). Taken together, these data suggested that RGS4 was a type 1 substrate of the N-end rule pathway in reticulocyte lysate.

At longer reaction times (either 2 or 4 h) the effect of type 1 N-end rule inhibitors on the relative level of RGS4 was not observed (Fig. 3, *A* and *C*, and Fig. 4). By contrast, the band of a model N-end rule substrate such as Arg-nsP4* was stronger in the presence of Arg-β-Ala inhibitor at all reaction times, including 2 and 4 h (Fig. 1*B* and Fig. 4). This difference in the patterns of inhibition of RGS4 and Arg-nsP4* degradation could be explained if, in contrast to Arg-nsP4*, there were two pools of the newly formed RGS4 molecules, one of which underwent rapid degradation, while the other remained long-lived and accumulated during the chase, after ~30 min of incubation. The fraction of short-lived RGS4 would be strongly increased by the type 1 inhibitors Arg-β-Ala or Lys-Ala at the early times of incubation (before ~1 h), but over longer times (from 1 to 4 h) this degradation-susceptible form of RGS4 would still be degraded in the lysate, presumably because of incomplete inhibition of the N-end rule pathway by dipeptides, in addition to gradual destruction of the dipeptides by peptidases in the lysate. Thus, at reaction times of 2 h or longer, only the subset of long-lived RGS4 molecules remained.

The fraction of short-lived RGS4 comprised 75–80% of the total RGS4; the remainder of RGS4 molecules (15–25%) was long-lived. This estimate stemmed from the fact that at 30 min, the amount of RGS4 was 4 to 5 times higher in the presence of Arg-β-Ala than in its absence (Fig. 3*C*, Fig. 4, and data not shown). (That the effect of Arg-β-Ala did not result from its influence on the rates of transcription/translation in the lysate was indicated by control experiments with the long-lived Met-nsP4* (Fig. 1, *C* and *D*).) In contrast to RGS4, most Arg-nsP4* molecules were short-lived in the lysate. As a result, the relative amount of Arg-nsP4* was higher in the presence of Arg-β-Ala than in its absence *throughout* the 4-h incubation (Fig. 4), although the amount of Arg-nsP4* continued to decline after 1 h even in the presence of Arg-β-Ala (Fig. 1), for the reasons described above.

We also examined the effect of a proteasome inhibitor, MG115 (61), on the degradation of RGS4 and Arg-nsP4* in the lysate. MG115 at 10 μM markedly increased the levels of both RGS4 and Arg-nsP4* at the incubation times shorter than 1 h (Fig. 5, *A–D*). MG115 at 0.1 mM further increased the amounts of RGS4 and Arg-nsP4* at these incubation times (Fig. 5, *A–D*). In addition, the ladders of apparently multiubiquitylated RGS4 and Arg-nsP4* were enhanced in the presence of MG115 (Fig. 5, *A* and *C*), in contrast to the effect of N-end rule inhibitors, which decreased the relative level of multi-Ub ladders (Fig. 3*B*). These patterns were consistent with MG115 inhibiting post-ubiquitylation steps of the proteasome-mediated degradation of RGS4 and Arg-nsP4*, in contrast to the action of dipeptides, which inhibited pre-ubiquitylation steps (25, 61).

Similarly to the effects of dipeptide inhibitors (Fig. 1), the proteasome level inhibition by MG115 was also incomplete and resulted in a gradual decrease of the RGS4 and Arg-nsP4* levels after 30 min of incubation (Fig. 5). Note that this decrease led to the nearly complete disappearance of Arg-nsP4* by 4 h of incubation, whereas in the case of RGS4 the decrease stopped when 20–25% of RGS4 still remained in the lysate (Fig. 5, *B* and *D*). These data provided independent evidence for the existence of a degradation-resistant fraction of RGS4. In summary, we identified mouse RGS4 as a type 1 substrate of the N-end rule pathway in the reticulocyte lysate.

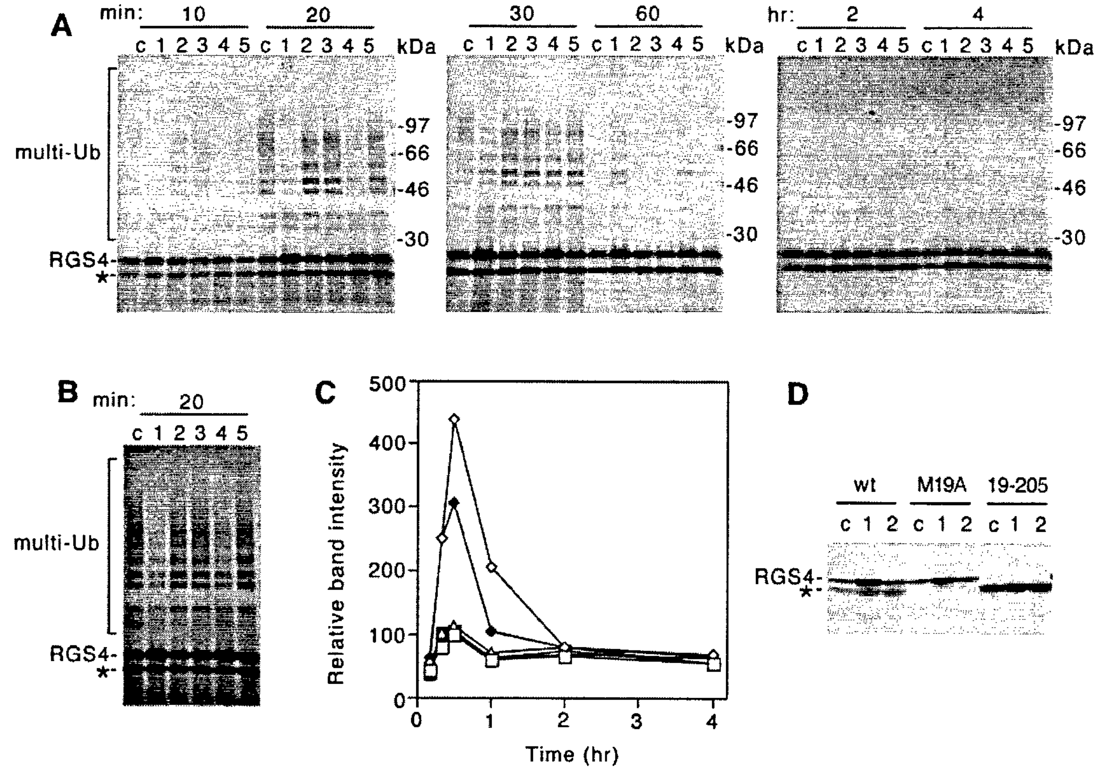


FIG. 3. **Type 1 dipeptides Arg-β-Ala and Lys-Ala inhibit degradation and ubiquitylation of RGS4 in reticulocyte lysate.** *A,* the plasmid pcDNA3-RGS4 (see "Experimental Procedures") was expressed in the transcription-translation reticulocyte lysate in the presence of [^{35}S]methionine and either in the absence of dipeptides (*lanes c*) or in the presence of 1 mM Arg-β-Ala (*lane 1*), 1 mM Trp-Ala (*lane 2*), 1 mM Ala-Lys (*lane 3*), 1 mM Lys-Ala (*lane 4*), and 1 mM Ala-Arg (*lane 5*). Dipeptide-containing samples also contained 0.15 mM bestatin. The reactions were carried out for 10, 20, 30, 60, 120, and 240 min, followed by SDS-PAGE, autoradiography, and quantitation. The positions of mouse RGS4 and the multi-Ub-containing species are indicated on the *left*. An *asterisk* marks the N-terminally truncated RGS4 derivatives, which were produced through alternative translational initiation from the internal start codon at the Met-19 position of RGS4. The positions of molecular mass markers are indicated on the *right*. *B,* autoradiographic overexposure of the 20-min part of *B* to visualize the ladders of multi-Ub-containing species. *C,* relative amounts of ^{35}S-RGS4 in *A* were determined using PhosphorImager and were plotted as a function of incubation time. The amount of RGS4 in the reaction without dipeptides at 30 min was assigned the value of 100. Control reactions were as follows: in the absence of dipeptides (□); in the presence of Arg-β-Ala (◇); in the presence of Trp-Ala (Δ); in the presence of Ala-Lys (■); in the presence of Lys-Ala (♦); and in the presence of Ala-Arg (▲). *D,* plasmids encoding either wild type RGS4 (*wt*), or RGS4$_{M19A}$, or RGS4$_{19-205}$ were expressed in the lysate either in the absence of dipeptides (*lane* c), or in the presence of 1 mM Arg-β-Ala and 0.15 mM bestatin (*lane 1*), or in the presence of 1 mM Trp-Ala and 0.15 mM bestatin (*lane 2*).

The Degradation Signal of RGS4 Is a Cysteine-based N-degron—The N-terminally truncated RGS4, produced through translation initiation at the second (Met-19) start codon of the *RGS4* ORF (see above), was not a substrate of the N-end rule pathway, because degradation of this RGS4$_{19-205}$ derivative in the lysate was not selectively decreased by dipeptide inhibitors of the N-end rule pathway (Fig. 3*A* and data not shown). This finding strongly suggested that the UBR1-dependent degron of the full-length RGS4 was located near its N terminus. The deduced N-terminal sequence of RGS4 is Met-Cys-Lys-Gly- (62). Extensive evidence indicates that cytosolic methionine aminopeptidases (MetAPs), a class of proteases present in all organisms, cleave off N-terminal Met from proteins and short peptides depending largely on the identity of the next residue, which becomes N-terminal after the cleavage. In particular, MetAPs rapidly (cotranslationally) cleave Met off N-terminal sequences beginning with Met-Cys (63–65). Cys is a stabilizing residue in prokaryotes and in the yeast *S. cerevisiae* but a secondary destabilizing residue in the metazoan N-end rule pathway (25, 26). Specifically, the destabilizing activity of N-terminal Cys is tRNA-dependent, strongly suggesting that an Arg-tRNA-protein transferase (R-transferase) conjugates Arg, a primary destabilizing residue, to the N-terminal Cys of an N-end rule substrate (25). This, and the fact that the degradation of RGS4 was inhibited by type 1 but not by type 2 dipeptides (Fig. 3), suggested the following model: the N-terminal Met of newly formed RGS4 is removed by MetAPs; the resulting N-terminal Cys is arginylated by R-transferase, and the arginylated RGS4 is targeted for degradation by the *UBR1*-encoded E3α and the rest of the N-end rule pathway.

One prediction of this model was that a significant fraction of RGS4 in the reticulocyte lysate should contain the N-terminal sequence Arg-Cys-Lys-Gly-, as distinguished from either Cys-Lys-Gly- or the encoded sequence Met-Cys-Lys-Gly-. Another prediction was that a mutational replacement of Cys-2 with a stabilizing residue that still allows the efficient removal of N-terminal Met should make the resulting RGS4 variant long-lived in the lysate. We tested both predictions.

Since the conventional N-terminal sequencing of RGS4 produced in the lysate would have required a major scale up of the reaction protocol, we employed radiochemical sequencing. A plasmid encoding RGS4$_{M19A}$ fused to glutathione transferase (GST) was constructed. The RGS4$_{M19A}$-GST protein was expressed in the lysate and was found to be degraded by the N-end rule pathway indistinguishably from the unmodified RGS4 (data not shown). The M19A mutation was introduced to preclude the formation of N-terminally truncated RGS4 variant (Fig. 3, *A* and *D*). RGS4$_{M19A}$-GST was synthesized in the lysate in the presence of Arg-β-Ala (type 1 inhibitor of the N-end rule pathway) and either [^{3}H]arginine, [^{3}H]lysine, or [^{3}H]leucine. The resulting ^{3}H-labeled RGS4$_{M19A}$-GST proteins were purified using glutathione-Sepharose affinity chromatography and SDS-PAGE. The band of RGS4$_{M19A}$-GST was subjected to automated Edman degradation, and ^{3}H released in each cycle was determined.

RGS4$_{M19A}$-GST labeled with [^{3}H]arginine yielded a major peak of ^{3}H in the first Edman cycle, followed by a second peak in the cycle 14, consistent with Arg at positions 1 and 14 (Fig. 6*A*). Whereas Arg at position 1 would have to be conjugated to RGS4 posttranslationally, as predicted by the above model, the first encoded Arg residue of RGS4 was, in fact, located at the encoded position 14 of arginylated RGS4 (in the sequence frame that begins with Arg-Cys-Lys-Gly-) (Fig. 6*A*). This experiment was carried out twice, with two independent preparations of [^{3}H]RGS4$_{M19A}$-GST, and gave the same result. RGS4$_{M19A}$-GST labeled with [^{3}H]lysine yielded virtually no at position 1 but produced a major peak at position 3 and elevations at positions 17 and 20, consistent with the known positions of three Lys residues in the N-terminal sequence of arginylated RGS4 (Fig. 6*B*). RGS4$_{M19A}$-GST labeled with [^{3}H]leucine yielded no ^{3}H in the cycle 1 but produced peaks in the cycles 5 and 8, once again consistent with the known positions of two Leu residues in the N-terminal sequence of arginylated RGS4 (Fig. 6*C*). Taken together, these independent sets of radiochemical sequencing data indicated that the bulk of RGS4 produced in reticulocyte lysate indeed bore the N-terminal sequence Arg-Cys-Lys-Gly-, with the Arg residue of this sequence not encoded by *RGS4* mRNA.

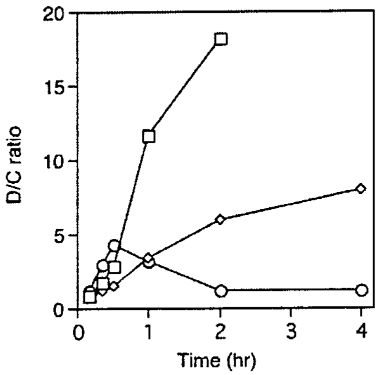


FIG. 4. **The relative levels of RGS4 and engineered N-end rule substrates in the presence of cognate dipeptide inhibitors in reticulocyte lysate.** ○, the ratios, denoted as dipeptides/controls (*D*/*C*), of the relative amounts of RGS4 in the presence *versus* the absence of 1 mM Arg-β-Ala, as a function of time in the lysate. These ratios were calculated from the data in Fig. 3*C*. ◇, the analogous ratio curve but for Tyr-nsP4* in the presence of 1 mM Trp-Ala (primary data not shown). □, the analogous ratio curve but for Arg-nsP4* in the presence of 1 mM Arg-β-Ala (calculated from the data in Fig. 1*B*; the 4-h ratio could not be determined reliably because of too low ^{35}S in the band of Arg-nsP4* in the absence of dipeptides). Note the monotonous increase of the dipeptides/controls ratio for engineered N-end rule substrate but not for RGS4.

To verify the second prediction of the model, we constructed several mutant alleles of RGS4 and expressed them in the lysate in either the absence or presence of N-end rule inhibitors (Fig. 7). It was found that the replacement of Cys-2 with either Gly, Val, or Ala residues completely stabilized the resulting RGS4$_{C2G}$, RGS$_{C2V}$, and RGS4$_{C2A}$ in the lysate (Fig. 7), in agreement with the model's prediction. We also converted Lys-3 of RGS4 into either Ser or Arg. Interestingly, whereas the Lys → Ser replacement completely stabilized the resulting RGS4$_{K3S}$ variant, the Lys → Arg replacement had no effect on the degradation of RGS4$_{K3R}$ (Fig. 7). Thus, although Lys-3 is not required for ubiquitylation of RGS4 (since it could be replaced with the non-ubiquitylatable Arg), the presence of a basic residue (either Lys or Arg) at position 3 (position 2 after the removal of N-terminal Met) is required for the RGS4 degradation by the N-end rule pathway.

The fact that RGS4$_{K3S}$, which bears N-terminal Cys and differs from RGS4 exclusively at position 3 (Lys → Ser replacement), was stable in the lysate (Fig. 7) indicated that the presence of N-terminal Cys was not sufficient for making a protein an N-end rule substrate. To explore this issue with other natural proteins, we expressed in the lysate the mouse protein Gβ$_{5L}$, a member of the family of β subunits of heterotrimeric G proteins (66). The predicted N-terminal sequence of Gβ$_{5L}$ is (Met)-Cys-Asp-Gln-Thr-. We found that Arg-β-Ala, the type 1 inhibitor of the N-end rule pathway, had no effect on the

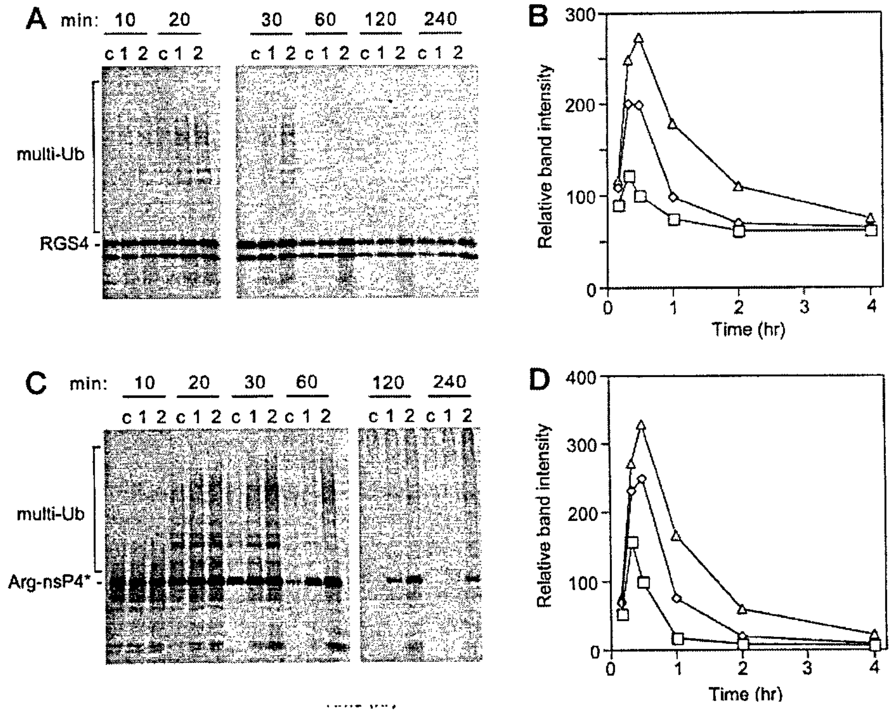


FIG. 5. **Proteasome inhibitor MG115 decreases degradation of RGS4 and Arg-nsP4* in reticulocyte lysate.** RGS4 (*A*) and Arg-nsP4* (*C*) were expressed in the transcription-translation reticulocyte lysate in the presence of [^{35}S]methionine and either in the absence of MG115 (Me$_2$SO solvent alone; *lanes c*) or in the presence of 10 μM (*lane 1*) and 100 μM (*lane 2*) MG115. The reactions were carried out for 10, 20, 30, 60, 120, and 240 min, followed by SDS-PAGE, autoradiography, and quantitation. The positions of RGS4 and Arg-nsP4* bands and of the ladder of multi-Ub-containing species are indicated on the *left*. *B*, relative amounts of ^{35}S-RGS4 in *A* were determined using PhosphorImager and were plotted as a function of incubation time. The amount of RGS4 in the reaction without MG115 at 30 min was assigned the value of 100. □, control reaction, in the absence of MG115; ◇, in the presence of 10 μM MG115; Δ, in the presence of 100 μM MG115. *D*, the same as in *B* but for Arg-nsP4*.

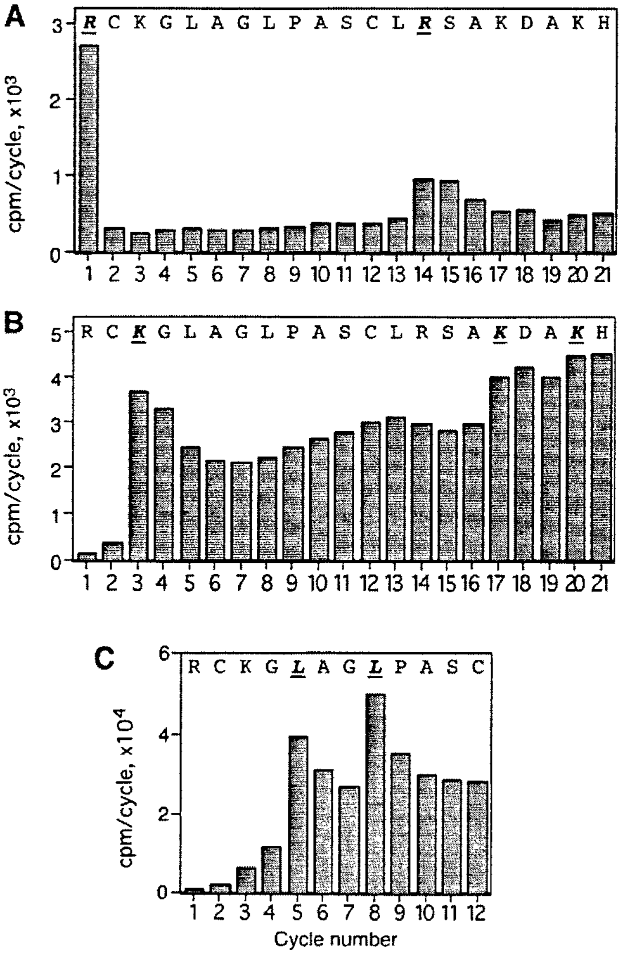


FIG. 6. **Radiochemical N-terminal sequencing of RGS4 produced in the reticulocyte lysate.** RGS4$_{M19A}$-GST was synthesized in the transcription-translation reticulocyte lysate for 30 min in the presence of 1 mM Arg-β–Ala and one of the ^{3}H-containing amino acids: [^{3}H]arginine (*A*), [^{3}H]lysine (*B*), or [^{3}H]leucine (*C*). The resulting ^{3}H-labeled RGS4$_{M19A}$-GST was purified by affinity chromatography and SDS-PAGE, followed by radiochemical N-terminal sequencing through Edman degradation, as described under "Experimental Procedures." Plotted are the amounts of ^{3}H (cpm) recovered in each Edman cycle. The inferred N-terminal sequence of RGS4$_{M19A}$-GST (see "Results") is shown at the *top* of each panel, with the expected ^{3}H residues highlighted in *bold*.

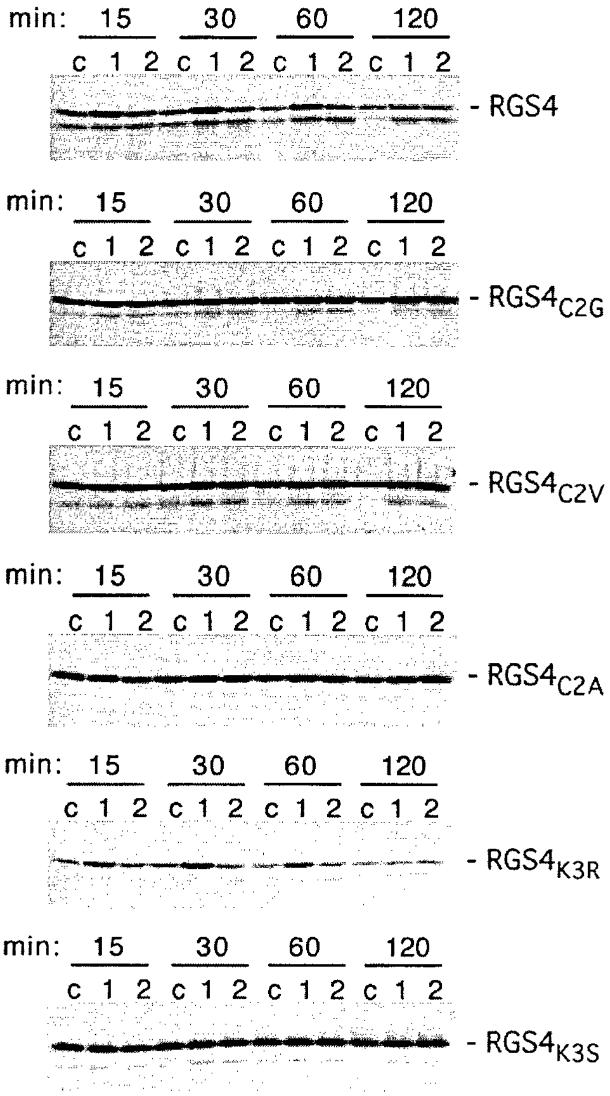


FIG. 7. **Single residue substitutions at the N terminus of RGS4 abolish its degradation in reticulocyte lysate.** Wild type mouse RGS4 and the indicated single residue derivatives of RGS4p were expressed in the transcription-translation reticulocyte lysate in the presence of [^{35}S]methionine and either in the absence of N-end rule inhibitors (*lane c*), or in the presence of 1 mM Arg-β-Ala (*lane 2*), or in the presence of 1 mM Trp-Ala (*lane 2*). Dipeptide-containing samples also contained 0.15 mM bestatin. The reactions were carried out for 15, 30, 60, and 120 min, followed by SDS-PAGE and autoradiography.

pattern of accumulation of Gβ_{5L} in reticulocyte lysate (data not shown).

The encoded N-terminal sequence of RGS4 is similar to those of RGS5 and RGS16; the similarities include Cys-2 and a basic residue at position 3 (55, 67). To test whether RGS16 was an N-end rule substrate, we expressed RGS16 in reticulocyte lysate. As shown in Fig. 8*A*, the relative amount of RGS16 was indeed significantly higher in the presence of Arg-β-Ala (type 1 inhibitor) than either in the presence of Trp-Ala (type 2 inhibitor) or in the absence of dipeptides. The temporal pattern of the inhibitor-produced increase in the RGS16 concentration was similar to that described for RGS4 in Fig. 3*B* (note the 30- and 60-min points in Fig. 8*A*). The effect of Arg-β-Ala on the accumulation of RGS16 (Fig. 8*A*), although significant, was smaller than its effect on the accumulation of RGS4 (Fig. 3*A*). Longer autoradiographic exposures showed the ladders of multi-Ub chains, apparently conjugated to RGS16 (Fig. 8*B*). Consistent with the weaker effect of Arg-β-Ala, the relative level of RGS16-specific multi-Ub chains (in the absence of inhibitors) was lower than that of RGS4-specific multi-Ub chains (Figs. 3*A* and 8*B* and data not shown). Finally, the ubiquitylation of RGS16 was significantly delayed in the presence of Arg-β-Ala (type 1 inhibitor) but not in the presence of Trp-Ala (type 2 inhibitor) (Fig. 8*B*). Taken together, these data (Fig. 8) indicated that RGS16 was also a substrate of the N-end rule pathway in reticulocyte lysate.

RGS4 Is Metabolically Unstable in Vivo—A mouse RGS4-expressing plasmid was transiently transfected into mouse L cells, and the metabolic stability of RGS4 was measured in a pulse-chase assay, using an antibody to RGS4. The transiently expressed RGS4 was found to be short-lived in L cells, with the half-life of 40–50 min (Fig. 9, *A* and *B*). When L cells were treated with the proteasome inhibitor MG132, both before and during pulse-chase, RGS4 was significantly stabilized (Fig. 9, *A* and *B*). Thus, similarly to the results with reticulocyte lysate, the *in vivo* degradation of RGS4 was proteasome-dependent.

To verify whether the degradation of RGS4 *in vivo* required its Cys-based N-degron, L cells were transfected with the plasmids expressing the RGS4$_{\mathrm{C2G}}$ and RGS4$_{\mathrm{C2V}}$ mutants, and the *in vivo* decay curves of these proteins were compared with that of wild type RGS4. It was found that these variants of RGS4, which were completely stable in reticulocyte lysate (Fig. 7), were degraded in L cells similarly to wild type RGS4 (data not shown).

DISCUSSION

We searched for substrates of the mammalian N-end rule pathway by testing some of the previously known candidates and also by employing a modification of the sib selection-based *in vitro* screen (42, 43) in a transcription-translation reticulocyte lysate. Two N-end rule substrates thus identified were mouse RGS4 and RGS16. These proteins are members of the RGS family of GTPase-activating proteins (GAPs) that down-regulate specific G proteins (44–50, 67). We report the following results.

1) The *in vitro* activated, proteolytically processed larger (80 kDa) subunit of the cysteine protease m-calpain was previously shown to bear N-terminal Lys (59), a type 1-destabilizing residue in the N-end rule (10). We expressed this subunit, termed Lys-mCL, in the reticulocyte lysate as a Ub-Lys-mCL fusion. The resulting Lys-mCL was not a substrate of the N-end rule pathway, despite the presence of a destabilizing N-terminal residue. Since an N-degron is a bipartite signal (12, 13, 15), the absence of active N-degron from Lys-mCL could be due to the absence of a targetable internal Lys residue, the second determinant of an N-degron. Another possibility is a steric hindrance in the binding of UBR1 to the N-terminal lysine of Lys-mCL. Thus, a destabilizing N-terminal residue is not the sole essential determinant of an active N-degron, as demonstrated previously with engineered N-end rule substrates (12).

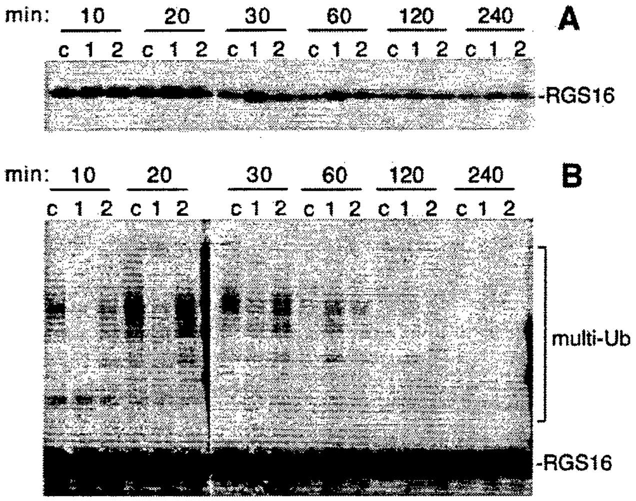

min: 10 20 30 60 120 240
A
c 1 2 c 1 2 c 1 2 c 1 2 c 1 2 c 1 2
-RGS16
min: 10 20 30 60 120 240
B
c 1 2 c 1 2 c 1 2 c 1 2 c 1 2 c 1 2
multi-Ub
-RGS16

FIG. 8. **Type 1 N-end rule inhibitor decreases degradation and ubiquitylation of RGS16 in reticulocyte lysate.** *A*, RGS16 was expressed in the transcription-translation reticulocyte lysate in the presence of [^{3}H]leucine and either in the absence of N-end rule inhibitors (*lane c*) or in the presence of 1 mM Arg-β-Ala (*lane 1*) or 1 mM Trp-Ala (*lane 2*). Dipeptide-containing samples also contained 0.15 mM bestatin. The reactions were carried out for 10, 20, 30, 60, 120, and 240 min, followed by SDS-PAGE and fluorography. *B*, the same as in *A*, but with autoradiographic overexposure to visualize the ladders of multi-Ub-containing species.

2) A major fraction (75–85%) of mouse RGS4 synthesized in reticulocyte lysate was rapidly degraded by a pathway that was proteasome-dependent and apparently also Ub-dependent. The other 15–25% of the newly made RGS4 was found to be stable in the lysate.

3) The degradation of RGS4 was inhibited by dipeptides bearing type 1-destabilizing N-terminal residues (Arg-β-Ala or Lys-Ala) but was unaffected by dipeptides bearing a type 2-destabilizing N-terminal residue (Trp-Ala) or a type 3 residue (Ala-Lys and Ala-Arg).

4) Radiochemical sequencing of RGS4 produced in reticulocyte lysate and labeled with either [^{3}H]arginine, [^{3}H]lysine, or [^{3}H]leucine revealed the presence of posttranslationally conjugated Arg at the N terminus of RGS4. The deduced N-terminal sequence of RGS4 was Met-Cys-Lys-Gly-. The observed/inferred sequence was Arg-Cys-Lys-Gly-. This result and, independently, the fact that degradation of RGS4 was inhibited by type 1 but not by type 2 dipeptides strongly suggested the following model: the N-terminal Met of newly formed RGS4 is removed by MetAPs; the resulting N-terminal Cys is arginylated by R-transferase; and the arginylated RGS4 is targeted for processive degradation by the *UBR1*-encoded E3α and the rest of the N-end rule pathway.

5) In agreement with a prediction of this model, the degradation of RGS4 in reticulocyte lysate was found to require Cys-2 residue, which becomes N-terminal after the MetAP-mediated cotranslational removal of Met. Specifically, the RGS4 mutants C2G, C2V, and C2A, in which Cys-2 was replaced with Gly, Val or Ala, were completely stable in the lysate. The former two residues are stabilizing in the N-end rule, and neither of the 3 residues is expected to interfere with the removal of N-terminal Met by MetAPs (64).

6) A Lys → Ser replacement at the encoded position 3 of RGS4 also completely stabilized the resulting RGS4$_{\mathrm{K3S}}$ protein. However, a Lys → Arg replacement at this position had no effect on the degradation of RGS4$_{\mathrm{K3R}}$. Thus, Lys-3 is not required for ubiquitylation of RGS4. However, the presence of a basic residue (either Lys or Arg) immediately after Cys is

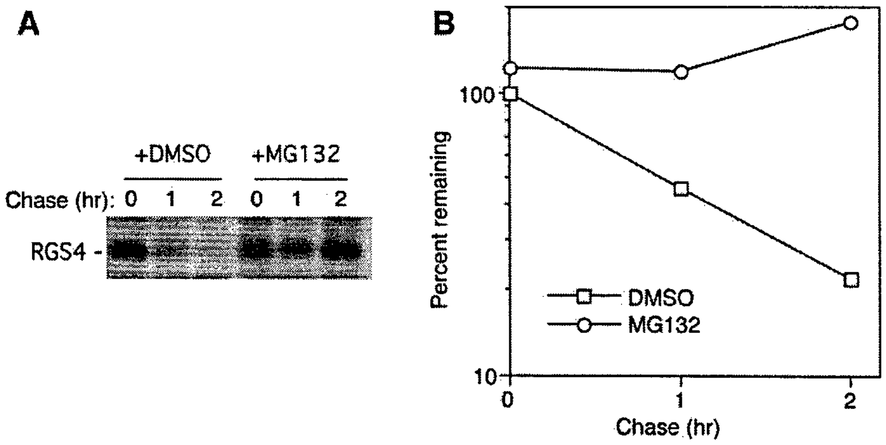


FIG. 9. ***In vivo* degradation of RGS4 in transiently transfected mouse L cells.** *A*, L cells were transfected with an RGS4-expressing plasmid (see "Experimental Procedures"). Cells were preincubated in methionine-free medium for 45 min and then labeled for 12 min with [^{35}S]methionine/cysteine and chased for 0, 1, or 2 h. Either 50 μM MG132 (from a stock solution in Me$_2$SO) or the equivalent amount of Me$_2$SO were present during both preincubation in methionine-free medium and the pulse-chase. Cell extracts were immunoprecipitated with anti-RGS4 antibody, followed by SDS-PAGE and autoradiography. *B*, the pattern in *A* was quantitated using PhosphorImager. The amount of ^{35}S in RGS4 at time 0 (end of the pulse) was taken as 100%. □, no proteasome inhibitor; circo, in the presence of 50 μM MG132.

essential for the RGS4 degradation by the N-end rule pathway. One possibility is that the requirement for a basic residue at this position reflects the substrate specificity of an uncharacterized R-transferase that arginylates the N-terminal Cys.

7) The mouse protein Gβ_{5L}, a member of the family of β subunits of heterotrimeric G proteins (66) that bore the (initial) N-terminal Met-Cys, was tested and found not to be an N-end rule substrate, similarly to the RGS4$_{K3S}$ mutant, which also bore the (initial) N-terminal sequence Met-Cys but was not degraded by the N-end rule pathway. Thus, the N-terminal Cys of wild type RGS4 is an essential determinant of its N-degron, but other N terminus-proximal residues are relevant as well.

8) Similarities among the N-terminal sequences of RGS4, RGS5, and RGS16 (67) suggested that RGS5 and RGS16 may also be N-end rule substrates. This prediction was tested, thus far, with RGS16, and was confirmed.

9) RGS4 was transiently expressed in mouse L cells and found to be an unstable protein, with the half-life of 40–50 min. The *in vivo* degradation of RGS4 was proteasome-dependent, as indicated by the nearly complete stabilization of RGS4 in the presence of proteasome inhibitor MG132. However, the targeting of RGS4 in this *in vivo* setting involved primarily a degron distinct from the Cys-based N-degron, because N-terminal variants of RGS4 such as RGS4$_{C2V}$ and RGS4$_{C2G}$, which were completely stable in reticulocyte lysate, remained unstable in L cells.

We consider the latter result first. 15–25% of the reticulocyte lysate-produced RGS4 was resistant to degradation by the N-end rule pathway (paragraph 2 above). An analogous protection of RGS4 against targeting by the N-end rule pathway in L cells might involve a much higher fraction of the newly formed RGS4. The mechanism of protection may be a modification of the N-terminal Cys, for example, its palmitoylation (45, 47, 62, 67) or acetylation (68) that would be expected to preclude the arginylation and degradation of RGS4 by the N-end rule pathway. Thus it is possible, indeed likely, that there is a kinetic competition among these reactions at the N-terminal Cys of a nascent RGS4 protein. In experiments to address this model, we added varying amounts of acetyl-CoA or palmitoyl-CoA (the substrates of N-terminal acetylases and palmitoyltransferases) to reticulocyte lysate; no effect of the added compounds on the degradation of newly formed RGS4 in the lysate was observed.[5] A second possibility is a spatial localization of RGS4 in L cells that protects it from degradation by the N-end rule pathway but leaves RGS4 still targetable by another proteasome-dependent pathway(s) that is inactive in reticulocytes. Yet another possible reason for the difference between the results with reticulocyte lysate *versus* L cells is that the cysteine branch of the N-end rule pathway may be cell type-specific; for example, a Cys-specific R-transferase is present in reticulocytes but might be expressed at a lower level in L cells. In recent experiments, RGS4 was expressed in *Xenopus laevis* oocytes, through microinjection of *RGS4* mRNA. It was found that, similarly to the results with reticulocyte extracts, RGS4 was degraded in oocytes by the cysteine branch of the N-end rule pathway.[6]

One difficulty in identifying physiological N-end rule substrates (as distinguished from those produced through the Ub fusion technique (10)) is caused by the fact that MetAPs remove N-terminal Met from newly formed proteins if, and only if, the second residue in a polypeptide chain is stabilizing in the yeast-type N-end rule. Specifically, the known MetAPs remove N-terminal Met if the second position residues are Gly, Val, Ala, Ser, Thr, Cys, or Pro (63–65). All of these residues are stabilizing in the yeast-type N-end rule (10). Therefore, a natural substrate of an N-end end rule pathway that bears an N-degron in, for instance, *S. cerevisiae* cells can be produced exclusively through cleavages (anywhere along the polypeptide chain) by proteases distinct from MetAPs. An example of N-end rule substrate of this class is the *S. cerevisiae* protein SCC1, a subunit of the cohesin complex that holds together sister chromatids. In a reaction that requires ESP1, the 566-residue SCC1 is cleaved (at the time of sister chromatid separation) at positions 180 and/or 268, resulting in fragments that bear N-terminal Arg, a type 1-destabilizing residue (69). These fragments were found to be degraded by the N-end rule pathway *in vivo.*[7]

The situation in metazoans such as mammals is similar to the one in yeast, in that the yeast-type destabilizing residues cannot be exposed at the N termini of mammalian proteins through cleavages by the known MetAPs (10, 64). The difference here is that Cys, Ala, Ser, Thr, and Pro are stabilizing residues in yeast but destabilizing in mammalian cells (10, 25, 70). In contrast to the other destabilizing residues, the second position Cys, Ala, Ser, Thr, and Pro can be efficiently exposed at the N termini of newly formed proteins through the action of MetAPs. Of these five residues, N-terminal Cys is a special case; its destabilizing activity requires its posttranslational conjugation to Arg, a primary destabilizing residue recognized by the type 1 site of UBR1 (E3α) (10).

The finding of the dependence of destabilizing activity of N-terminal Cys on the presence of tRNA (25, 26), and the identification, in the present work, of Arg as the posttranslationally linked N-terminal residue of mouse RGS4, indicated the existence of an Arg-tRNA-protein transferase (R-transferase) that conjugates Arg to N-terminal Cys. This R-transferase is distinct from the two previously characterized mammalian R-transferases, both of which are encoded by the alternatively spliced *ATE1* gene (27). The two forms of mouse ATE1 R-transferase have different specific activities but apparently the same substrate specificity; similarly to their *S. cerevisiae* homolog, they arginylate N-terminal Asp and Glu but cannot arginylate N-terminal Cys (27).

The ~20 known proteins of the mammalian RGS family have in common the ~130-residue RGS domain that binds to Gα subunits and is responsible for the GAP function of RGS proteins (44, 45). As would be expected of pleiotropic regulators at key junctions of the cellular metabolism, the activity of RGS proteins is controlled at several levels, including regulation of their synthesis and localization (44). It is becoming increasingly clear that the (conditional) degradation of RGSs is yet another way in which the activity of these proteins is modulated in cells. Besides RGS4 and RGS16 of the present work, one other member of the family, RGS7, was recently shown to be unstable *in vivo* (71, 72). The normally short-lived RGS7 is stabilized through its interaction with polycystin, a *PKD1*-encoded protein involved in polycystic kidney disease (71). The degradation of RGS7 is also decreased upon exposure of cells to tumor necrosis factor α; the resulting increase in the concentration of RGS7 may contribute to sepsis-induced changes in the nervous system (72). The degradation signal(s) of RGS7 remains to be characterized.

Our major results are the findings that RGS4 and RGS16 (and possibly also RGS5) are substrates of the cysteine branch of the N-end rule pathway and that the Cys-based N-degron of these proteins functions through the posttranslational arginylation of N-terminal Cys by an unidentified, apparently Cys-specific R-transferase. RGS4 and RGS16 are the first physiological substrates of the N-end rule pathway that bear a secondary destabilizing N-terminal residue, which functions through its conjugation to Arg, a primary destabilizing residue.

---

[5] I. Davydov and A. Varshavsky, unpublished data.

[6] J. Sheng, I. Davydov, and A. Varshavsky, unpublished data.

[7] H. Rao and A. Varshavsky, unpublished data.

Specific functions of the metabolic instability of RGS4 and RGS16 remain to be understood. One route to these functions is through the cloning of Cys-specific R-transferase and construction of mouse strains that lack this enzyme and (therefore) the cysteine branch of the N-end rule pathway. It would also be important to define, in detail, Cys-proximal N-terminal sequences in proteins that promote the arginylation of N-terminal Cys. This information will facilitate the identification of Cys-specific N-end rule substrates in data bases of protein sequences.

Acknowledgments—We thank T. Rümenapf for providing Ub-X-nsP4-encoding plasmids, which were used as PCR templates in the present work, and for sharing unpublished data on the N-end rule-dependent degradation of truncated Tyr-nsP4 in reticulocyte lysate. We are grateful to G. Hathaway for expert assistance with the radiochemical sequencing of RGS4$_{M19A}$-GST and to Y. T. Kwon and N. Barteneva for the mouse brains used to prepare the cDNA library. We thank C. K. Chen and M. I. Simon for the gifts of pcDNA3-RGS16 and a Gβ_{5L}-coding DNA fragment; H. Sorimachi and K. Suzuki for the gifts of p94- and p94$_{C129A}$-encoding plasmids; and J. S. Elce for the gift of pET-80k. We gratefully acknowledge discussions of the paper and comments on the manuscript by E. Smirnova and members of the Varshavsky laboratory.

REFERENCES

1. Maniatis, T. (1999) *Genes Dev.* **13,** 505–510
2. Nasmyth, K. (1999) *Trends Biochem. Sci.* **24,** 98–104
3. Peters, J.-M., King, R. W., and Deshaies, R. J. (1998) in *Ubiquitin and the Biology of the Cell* (Peters, J.-M., Harris, J. R., and Finley, D., eds), pp. 345–387, Plenum Publishing Corp., New York
4. Hershko, A., and Ciechanover, A. (1998) *Annu. Rev. Biochem.* **76,** 425–479
5. Varshavsky, A. (1997) *Trends Biochem. Sci.* **22,** 383–387
6. Wickner, S., Maurizi, M. R., and Gottesman, S. (1999) *Science* **286,** 1888–1893
7. Varshavsky, A. (1991) *Cell* **64,** 13–15
8. Laney, J. D., and Hochstrasser, M. (1999) *Cell* **97,** 427–430
9. Bachmair, A., Finley, D., and Varshavsky, A. (1986) *Science* **234,** 179–186
10. Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 12142–12149
11. Kwon, Y. T., Reiss, Y., Fried, V. A., Hershko, A., Yoon, J. K., Gonda, D. K., Sangan, P., Copeland, N. G., Jenkins, N. A., and Varshavsky, A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 7898–7903
12. Bachmair, A., and Varshavsky, A. (1989) *Cell* **56,** 1019–1032
13. Johnson, E. S., Gonda, D. K., and Varshavsky, A. (1990) *Nature* **346,** 287–291
14. Pickart, C. M. (1997) *FASEB J.* **11,** 1055–1066
15. Suzuki, T., and Varshavsky, A. (1999) *EMBO J.* **18,** 6017–6026
16. Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K., and Varshavsky, A. (1989) *Science* **243,** 1576–1583
17. Hochstrasser, M. (1996) *Annu. Rev. Genet.* **30,** 405–439
18. Baumeister, W., Walz, J., Zühl, F., and Seemüller, E. (1998) *Cell* **92,** 367–380
19. DeMartino, G. N., and Slaughter, C. A. (1999) *J. Biol. Chem.* **274,** 22123–22126
20. Baker, R. T., and Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 12065–12074
21. Balzi, E., Choder, M., Chen, W., Varshavsky, A., and Goffeau, A. (1990) *J. Biol. Chem.* **265,** 7464–7471
22. Li, J., and Pickart, C. M. (1995) *Biochemistry* **34,** 15829–15837
23. Stewart, A. E., Arfin, S. M., and Bradshaw, R. A. (1995) *J. Biol. Chem.* **270,** 25–28
24. Grigoryev, S., Stewart, A. E., Kwon, Y. T., Arfin, S. M., Bradshaw, R. A., Jenkins, N. A., Copeland, N. G., and Varshavsky, A. (1996) *J. Biol. Chem.* **271,** 28521–28532
25. Gonda, D. K., Bachmair, A., Wünning, I., Tobias, J. W., Lane, W. S., and Varshavsky, A. (1989) *J. Biol. Chem.* **264,** 16700–16712
26. Davydov, I. V., Patra, D., and Varshavsky, A. (1998) *Arch. Biochem. Biophys.* **357,** 317–325
27. Kwon, Y. T., Kashina, A. S., and Varshavsky, A. (1999) *Mol. Cell. Biol.* **19,** 182–193
28. Reiss, Y., Kaim, D., and Hershko, A. (1988) *J. Biol. Chem.* **263,** 2693–2698
29. Baker, R. T., and Varshavsky, A. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 2374–2378
30. Byrd, C., Turner, G. C., and Varshavsky, A. (1998) *EMBO J.* **17,** 269–277
31. Turner, G. C., Du, F., and Varshavsky, A. (2000) *Nature* **405,** 579–583
32. Schauber, C., Chen, L., Tongaonkar, P., Vega, I., and Madura, K. (1998) *Genes Cells* **3,** 307–319
33. Lawson, T. G., Gronros, D. L., Evans, P. E., Bastien, M. C., Michalewich, K. M., Clark, J. K., Edmonds, J. H., Graber, K. H., Werner, J. A., Lurvey, B. A., and Cate, J. M. (1999) *J. Biol. Chem.* **274,** 9871–9880
34. deGroot, R. J., Rümenapf, T., Kuhn, R. J., and Strauss, J. H. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 8967–8971
35. Sijts, A. J., Pilip, I., and Pamer, E. G. (1997) *J. Biol. Chem.* **272,** 19261–19268
36. Obin, M., Mesco, E., Gong, X., Haas, A. L., Joseph, J., and Taylor, A. (1999) *J. Biol. Chem.* **274,** 11789–11795
37. Hondermarck, H., Sy, J., Bradshaw, R. A., and Arfin, S. M. (1992) *Biochem. Biophys. Res. Commun.* **30,** 280–288
38. Taban, C. H., Hondermarck, H., Bradshaw, R. A., and Boilly, B. (1996) *Experientia (Basel)* **52,** 865–870
39. Solomon, V., Baracos, V., Sarraf, P., and Goldberg, A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 12602–12607
40. Wilkinson, K., and Hochstrasser, M. (1998) in *Ubiquitin and the Biology of the Cell* (Peters, J.-M., Harris, J. R., and Finley, D., eds) pp. 99–125, Plenum Publishing Corp., New York
41. Baker, R. T. (1996) *Curr. Opin. Biotechnol.* **7,** 541–546
42. Lustig, K. D., Stukenberg, P. T., McGarry, T. J., King, R. W., Cryns, V. L., Mead, P. E., Zon, L. I., Yuan, J., and Kirschner, M. W. (1997) *Methods Enzymol.* **283,** 83–99
43. Stukenberg, P. T., Lustig, K. D., McGarry, T. J., King, R. W., Kuang, J., and Kirschner, M. W. (1997) *Curr. Biol.* **7,** 338–348
44. De Vries, L., and Farquhar, M. G. (1999) *Trends Cell Biol.* **9,** 138–144
45. Berman, D. M., and Gilman, A. G. (1998) *J. Biol. Chem.* **273,** 1269–1272
46. Heximer, S. P., Srinivasa, S. P., Bernstein, L. S., Bernard, J. L., Linder, M. E., Hepler, J. R., and Blumer, K. J. (1999) *J. Biol. Chem.* **274,** 34253–34259
47. Tu, Y., Popov, S., Slaughter, C., and Ross, E. M. (1999) *J. Biol. Chem.* **274,** 38260–38267
48. DiBello, P. R., Garrison, T. R., Apanovitch, D. M., Hoffman, G., Shuey, D. J., Mason, K., Cockett, M. I., and Dohlman, H. G. (1998) *J. Biol. Chem.* **273,** 5780–5784
49. Yan, Y., Chi, P. P., and Bourne, H. R. (1997) *J. Biol. Chem.* **272,** 11924–11927
50. Wieland, T., Chen, C. K., and Simon, M. I. (1997) *J. Biol. Chem.* **272,** 8853–8856
51. Lemm, J. A., Rümenapf, T., Strauss, E. G., Strauss, J. H., and Rice, C. M. (1994) *EMBO J.* **13,** 2295–2934
52. Elce, J., Hegadorn, C., Simon, J., and Arthur, J. S. C. (1997) *J. Biol. Chem.* **272,** 11268–11275
53. Sorimachi, H., Toyamasorimachi, N., Saido, T. C., Kawasaki, H., Sugita, H., Miyasaka, M., Arahata, K., Ishimura, S., and Suzuki, K. (1993) *J. Biol. Chem.* **268,** 10593–10605
54. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K. (eds) (1998) *Current Protocols in Molecular Biology*, Wiley-Interscience, New York
55. Chen, C.-K., Wieland, T., and Simon, M. I. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 12885–12889
56. Lévy, F., Johnsson, N., Rumenapf, T., and Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 4907–4912
57. Johnsson, N., and Varshavsky, A. (1994) *EMBO J.* **13,** 2686–2698
58. Sorimachi, H., Ishiura, S., and Suzuki, K. (1997) *Biochem. J.* **328,** 721–732
59. Brown, N., and Crawford, C. (1993) *FEBS Lett.* **322,** 65–68
60. Kinbara, K., Ishiura, S., Tomioka, S., Sorimachi, H., Jeong, S. Y., Amano, S., Kawasaki, H., Kolmerer, B., Kimura, S., Labeit, S., and Suzuki, K. (1998) *Biochem. J.* **335,** 589–596
61. Rock, K. L., Gramm, C., Rothstein, L., Clark, K., Stein, R., Dick, L., Hwang, D., and Goldberg, A. L. (1994) *Cell* **78,** 761–771
62. Druey, K. M., Blumer, K. J., Kang, V. H., and Kehrl, J. H. (1996) *Nature* **379,** 742–746
63. Moerschell, R. P., Hosokawa, Y., Tsunasawa, S., and Sherman, F. (1990) *J. Biol. Chem.* **265,** 19638–19643
64. Walker, K. W., and Bradshaw, R. A. (1999) *J. Biol. Chem.* **274,** 13403–13409
65. Turk, B. E., Griffith, E. C., Wolf, S., Biemann, C., Chang, Y.-H., and Liu, J. O. (1999) *Chem. Biol.* **6,** 823–833
66. Watson, A. J., Agaray, A. M., Slepak, V. Z., and Simon, M. I. (1996) *J. Biol. Chem.* **271,** 28154–28160
67. Srinivasa, S. P., Bernstein, L. S., Blumer, K. J., and Linder, M. E. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 5584–5589
68. Cook, R. K., Sheff, D. R., and Rubenstein, P. A. (1991) *J. Biol. Chem.* **266,** 16825–16833
69. Uhlmann, F., Lottspeich, F., and Nasmyth, K. (1999) *Nature* **400,** 37–42
70. Nishizawa, M., Furuno, N., Okazaki, K., Tanaka, H., Ogawa, Y., and Sagata, N. (1993) *EMBO J.* **12,** 4021–4027
71. Kim, E., Arnould, T., Sellin, L., Benzing, T., Comella, N., Kocher, O., Tsiokas, L., Sukhatme, V. P., and Walz, G. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 6371–6376
72. Benzing, T., Brandes, R., Sellin, L., Schermer, B., Lecker, S., Walz, G., and Kim, E. (1999) *Nat. Med.* **5,** 913–918
73. Stewart, A. (1995) *Trends in Genetics Nomenclature Guide*, Elsevier Science, Ltd., Cambridge, U.K.
74. Webb, E. C. (ed) (1992) *Enzyme Nomenclature 1992*, p.527, Academic Press, New York

[41] Ubiquitin Fusion Technique and Its Descendants

By ALEXANDER VARSHAVSKY

The ubiquitin (Ub) fusion technique was developed in 1985–1986, through experiments in which a segment of DNA encoding the 76-residue Ub was joined, in frame, to DNA encoding *Escherichia coli* β-galactosidase (βgal).[1,2] When the resulting protein fusion was expressed in the yeast *Saccharomyces cerevisiae* and detected by radiolabeling and immunoprecipitation with an anti-βgal antibody, only the moiety of βgal was observed, even if the labeling time was short enough to be comparable to the time (1–2 min) required for translation of the Ub-βgal open reading frame (ORF). It was found that in eukaryotic cells the Ub moiety of the fusion was rapidly cleaved off after the last residue of Ub (Fig. 1).[1] The proteases involved are called deubiquitylating[3] enzymes (DUBs) or Ub-specific processing proteases (UBPs).[4–7] A eukaryotic cell contains more than 10 distinct DUBs, all of which are highly specific for the Ub moiety. The *in vivo*

---

[1] A. Bachmair, D. Finley, and A. Varshavsky, *Science* **234**, 179 (1986).

[2] A. Varshavsky, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12142 (1996).

[3] Ubiquitin whose C-terminal (Gly-76) carboxyl group is covalently linked to another compound is called the *ubiquityl* moiety, the derivative terms being *ubiquitylation* and *ubiquitylated*. The term *Ub* refers to both free ubiquitin and the ubiquityl moiety. This nomenclature, which is also recommended by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology,[19] brings Ub-related terms in line with the standard chemical terminology.

[4] K. Wilkinson and M. Hochstrasser, *in* "Ubiquitin and the Biology of the Cell" (J.-M. Peters, J. R. Harris, and D. Finley, eds.). Plenum Press, New York, 1998.

[5] J. W. Tobias and A. Varshavsky, *J. Biol. Chem.* **266**, 12021 (1991).

[6] R. T. Baker, J. W. Tobias, and A. Varshavsky, *J. Biol. Chem.* **267**, 23364 (1992).

[7] C. A. Gilchrist, D. A. Gray, and R. T. Baker, *J. Biol. Chem.* **272**, 32280 (1997).

Copyright © 2000 by Academic Press
All rights of reproduction in any form reserved.

METHODS IN ENZYMOLOGY, VOL. 327
0076-6879/00 $30.00

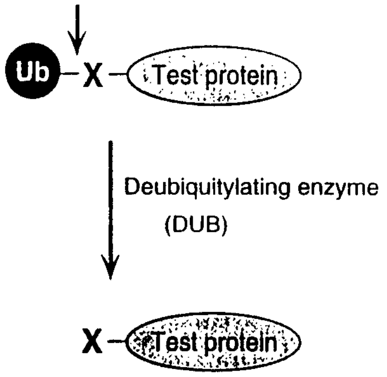


FIG. 1. The ubiquitin fusion technique. Linear fusions of Ub to other proteins are cleaved after the last residue of Ub by deubiquitylating enzymes (DUBs) (see text).[1,2]

cleavage at the Ub–polypeptide junction of a Ub fusion has been shown to be largely cotranslational.[8,9]

One physiological function of the cleavage reaction (Fig. 1) is to mediate the excision of Ub from its natural DNA-encoded fusions either to itself (poly-Ub)[10] or to specific ribosomal proteins.[11,12] Many of the DUB proteases that catalyze the cleavage of linear Ub fusions can also cleave Ub off its branched, posttranslationally formed conjugates, in which Ub is joined either to itself, as in a multi-Ub chain, or to other proteins.[4,13] A branched Ub–protein conjugate usually comprises a multi-Ub chain covalently linked to an internal lysine residue of a substrate protein. The ubiquitylated substrate is processively degraded by the 26S proteasome, an ATP-dependent multisubunit protease.[14–17] For reviews of the Ub system, see Refs. 18–22.

Another finding about the DUB-mediated cleavage reaction (Fig. 1)

[8] N. Johnsson and A. Varshavsky, *EMBO J.* **13**, 2686 (1994).

[9] G. C. Turner and A. Varshavsky, submitted (2000).

[10] D. Finley, E. Özkaynak, and A. Varshavsky, *Cell* **48**, 1035 (1987).

[11] K. L. Redman and M. Rechsteiner, *Nature* (*London*) **338**, 438 (1989).

[12] D. Finley, B. Bartel, and A. Varshavsky, *Nature* (*London*) **338**, 394 (1989).

[13] C. M. Pickart, *FASEB J.* **11**, 1055 (1997).

[14] O. Coux, K. Tanaka, and A. L. Goldberg, *Annu. Rev. Biochem.* **65**, 801 (1996).

[15] W. Baumeister, J. Walz, F. Zühl, and E. Seemüller, *Cell* **92**, 367 (1998).

[16] M. Rechsteiner, *in* "Ubiquitin and the Biology of the Cell" (J. M. Peters, J. R. Harris, and D. Finley, eds.), pp. 147–189. Plenum Press, New York, 1998.

[17] G. N. DeMartino and C. A. Slaughter, *J. Biol. Chem.* **274**, 22123 (1999).

[18] M. Hochstrasser, *Annu. Rev. Genet.* **30**, 405 (1996).

[19] A. Varshavsky, *Trends Biochem. Sci.* **22**, 383 (1997).

[20] A. Hershko and A. Ciechanover, *Annu. Rev. Biochem.* **76**, 425 (1998).

[21] T. Maniatis, *Genes Dev.* **13**, 505 (1999).

[22] L. Hicke, *Trends Cell Biol.* **9**, 107 (1999).

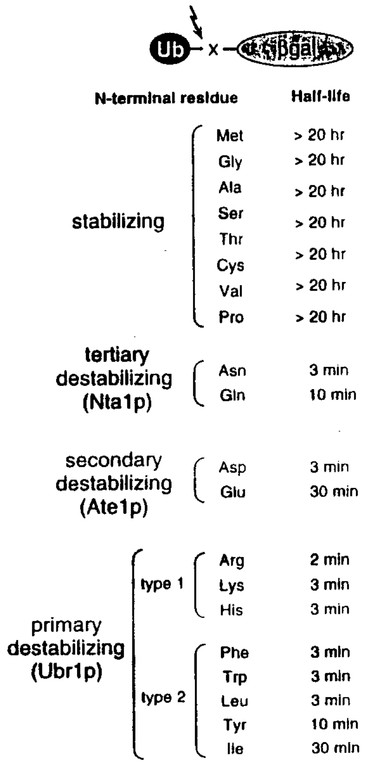

Ub
X
βgal
N-terminal residue
Half-life
stabilizing
Met > 20 hr
Gly > 20 hr
Ala > 20 hr
Ser > 20 hr
Thr
Cys > 20 hr
Val > 20 hr
Pro > 20 hr
tertiary destabilizing (Nta1p)
Asn 3 min
Gln 10 min
secondary destabilizing (Ate1p)
Asp 3 min
Glu 30 min
primary destabilizing (Ubr1p)
type 1
Arg 2 min
Lys 3 min
His 3 min
type 2
Phe 3 min
Trp 3 min
Leu 3 min
Tyr 10 min
Ile 30 min

FIG. 2. The N-end rule of the yeast *S. cerevisiae*.[2] Specific residues at the N terminus of a test protein such as βgal are produced by the Ub fusion technique (Fig. 1 and text). The *in vivo* half-lives of the corresponding *X*-βgal proteins are indicated on the right. Stabilizing N-terminal residues (Met, Gly, Ala, Ser, Thr, Cys, Val, and Pro) are not recognized by Ubr1p (N-recognin), the E3 component of the N-end rule pathway. Primary destabilizing N-terminal residues (Arg, Lys, His, Phe, Trp, Leu, Tyr, and Ile) are directly bound by either type 1 or type 2 substrate-binding sites of Ubr1p. Secondary destabilizing N-terminal residues are arginylated by the *ATE1*-encoded Arg-tRNA-protein transferase (R-transferase), yielding the N-terminal Arg, a primary destabilizing residue. Tertiary destabilizing N-terminal residues Asn and Gln are deamidated by the *NTA1*-encoded N-terminal amidohydrolase (Nt-amidase), yielding the secondary destabilizing residues Asp and Glu, respectively. The N-end rule of mammalian cells is similar but contains fewer stabilizing residues.[2]

led to the discovery of the N-end rule, a relation between the *in vivo* half-life of a protein and the identity of its N-terminal residue (Fig. 2).[1] First, it was shown that the cleavage of a Ub–*X*–polypeptide fusion after the last residue of Ub takes place regardless of the identity of a residue *X* at the C-terminal side of the cleavage site, proline being the single exception. By allowing a bypass of the "normal" N-terminal processing of a newly formed protein, this result yielded an *in vivo* method for placing different residues at the N termini of otherwise identical proteins. Second, it was found that the *in vivo* half-lives of the resulting test proteins were determined by the identities of their N-terminal residues, a relation referred to as the N-end rule (Fig. 2).[1] The N-end rule pathway, which targets the N terminus-specific degradation signals, called the N-degrons, is one pathway of the Ub system. For a review and work on the N-end rule pathway, see Refs. 2 and 23–31.

The Ub fusion technique (Figs. 1 and 2) remains the method of choice for producing, *in vivo*, the desired N-terminal residue in a protein of interest. Owing to the constraints of the genetic code, nascent proteins bear N-terminal methionine (formyl-Met in prokaryotes). The known methionine aminopeptidases (MAPs), which remove N-terminal Met, do so only if the residue to be exposed is stabilizing according to the yeast-type N-end rule.[2,32] In other words, MAPs do not cleave off N-terminal methionine if it is followed by any of the 12 destabilizing residues (Fig. 2). The Ub-specific DUB proteases are free of this constraint, except when the residue *X* of a Ub–*X*–polypeptide is proline, in which case the cleavage still takes place but at a much lower rate.[1,33] More recently, a specific DUB was identified that can efficiently cleave at the Ub–proline junction.[7]

The Ub fusions can be deubiquitylated *in vitro* as well.[25,34,35] The high activity and specificity of DUBs should make them the reagents of choice for applications that involve, for example, the removal of affinity tags from overexpressed and purified proteins. Unfortunately, there are no commercially available DUBs at present, in part because of difficulties encountered in purifying and stabilizing large DUBs such as *S. cerevisiae* Ubp1p, and also because Proteinix (Rockville, MD), a company that has held the licenses for Ub fusion patents over the last decade, has not commercialized this technology.

Another major application of the Ub fusion technique resulted from the observations that expression of a protein as a Ub fusion can dramatically

---

[23] C. Byrd, G. C. Turner, and A. Varshavsky, *EMBO J.* **17**, 269 (1998).

[24] Y. T. Kwon, Y. Reiss, V. A. Fried, A. Hershko, J. K. Yoon, D. K. Gonda, P. Sangan, N. G. Copeland, N. A. Jenkins, and A. Varshavsky, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7898 (1998).

[25] I. V. Davydov, D. Patra, and A. Varshavsky, *Arch. Biochem. Biophys.* **357**, 317 (1998).

[26] Y. T. Kwon, A. S. Kashina, and A. Varshavsky, *Mol. Cell. Biol.* **19**, 182 (1999).

[27] Y. T. Kwon, F. Lévy, and A. Varshavsky, *J. Biol. Chem.* **274**, 18135 (1999).

[28] F. Lévy, J. A. Johnston, and A. Varshavsky, *Eur. J. Biochem.* **259**, 244 (1999).

[29] T. Suzuki and A. Varshavsky, *EMBO J.* **18**, 101 (1999).

[30] Y. Xie and A. Varshavsky, *Curr. Genet.* **36**, 113 (1999).

[31] P. O. Falnes and S. Olsnes, *EMBO J.* **17**, 615 (1998).

[32] R. A. Bradshaw, W. W. Brickey, and K. W. Walker, *Trends Biochem. Sci.* **23**, 263 (1998).

[33] E. S. Johnson, B. W. Bartel, and A. Varshavsky, *EMBO J.* **11**, 497 (1992).

[34] D. K. Gonda, A. Bachmair, I. Wünning, J. W. Tobias, W. S. Lane, and A. Varshavsky, *J. Biol. Chem.* **264**, 16700 (1989).

[35] R. T. Baker, *Curr. Opin. Biotechnol.* **7**, 541 (1996).

augment the yield of the protein.[36-39] The yield enhancement effect of Ub was observed with short peptides as well.[40,41] This and other applications of Ub fusions are described below, with references to the original articles and specific constructs.

Production and Uses of N-Degrons

An N-degron comprises the destabilizing N-terminal residue of a protein and an internal lysine residue.[2,29,42,43] A set of N-degrons containing different N-terminal residues that are destabilizing in a given cell defines the N-end rule of the cell.[2] The lysine determinant of an N-degron is the site of formation of a substrate-linked multi-Ub chain.[13,18,44] A way to produce an N-degron in a protein of interest is to express the protein as a Ub fusion in which the junctional residue (which becomes N-terminal on removal of the Ub moiety) is destabilizing (Fig. 2). An appropriately positioned internal lysine residue (or residues) is the second essential determinant of N-degron. Many natural proteins lack such "targetable" lysines, and therefore would remain long-lived even if their N-terminal residue were replaced by a destabilizing residue. One way to bypass this difficulty is to link a protein of interest to a relatively short (<50 residues) portable N-degron that contains both an N-terminal destabilizing residue (produced through a Ub fusion) and a requisite lysine residue(s). The earliest portable N-degron of this kind is still among the strongest known (Fig. 3B).[1,29,42] It was found, using the new strategy of a screen in the sequence space of just two amino acids, lysine and asparagine, that certain sequences containing exclusively lysines and asparagines can function *in vivo* as highly effective N-degrons.[29] The portability and modular organization of N-degrons make possible a variety of applications whose common feature is the conferring of a constitutive or conditional metabolic instability on a protein of interest.

---

[36] T. R. Butt, S. Jonnalagadda, B. P. Monia, E. J. Sternberg, J. A. Marsh, J. M. Stadel, D. J. Ecker, and S. T. Crooke, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2540 (1989).

[37] D. J. Ecker, J. M. Stadel, T. R. Butt, J. A. Marsh, B. P. Monia, D. A. Powers, J. A. Gorman, P. E. Clark, F. Warren, and A. Shatzman, *J. Biol. Chem.* **264**, 7715 (1989).

[38] P. Mak, D. P. McDonnell, N. L. Weigel, W. T. Schrader, and B. W. O'Malley, *J. Biol. Chem.* **264**, 21613 (1989).

[39] R. T. Baker, S. A. Smith, R. Marano, J. McKee, and P. G. Board, *J. Biol. Chem.* **269**, 25381 (1994).

[40] Y. Yoo, K. Rote, and M. Rechsteiner, *J. Biol. Chem.* **264**, 17078 (1989).

[41] A. Pilon, P. Yost, T. E. Chase, G. Lohnas, T. Burkett, S. Roberts, and W. E. Bentley, *Biotechnol. Prog.* **13**, 374 (1997).

[42] A. Bachmair and A. Varshavsky, *Cell* **56**, 1019 (1989).

[43] C. P. Hill, N. L. Johnston, and R. E. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4136 (1993).

[44] V. Chau, J. W. Tobias, A. Bachmair, D. Marriott, D. J. Ecker, D. K. Gonda, and A. Varshavsky, *Science* **243**, 1576 (1989).

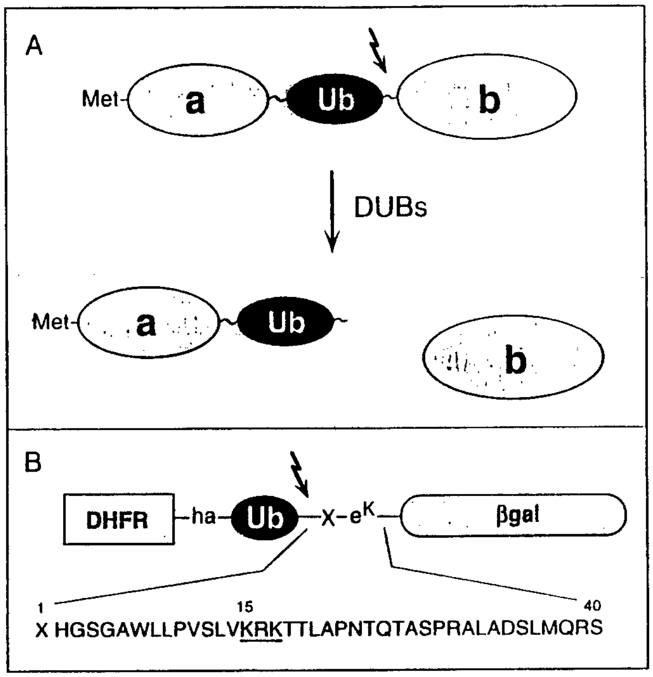


FIG. 3. **The UPR (ubiquitin/protein/reference) technique.**[79] (A) A tripartite fusion containing **a**, the reference protein moiety whose C terminus is linked, via a spacer peptide, to the Ub moiety. The C terminus of Ub is linked to **b**, a protein of interest. *In vivo*, this tripartite fusion is cotranslationally cleaved[9] by deubiquitylating enzymes (DUBs) at the Ub–**b** junction, yielding equimolar amounts of the unmodified protein **b** and **a**–Ub, the reference protein **a** bearing a C-terminal Ub moiety. If **a**–Ub is long-lived, a measurement of the ratio of **a**-Ub to **b** as a function of time or at steady state yields, respectively, the *in vivo* decay curve or the relative metabolic stability of protein **b**.[29,79] (B) Example of a specific UPR-type Ub fusion.[29] This fusion contains the following elements: DHFRha, a mouse dihydrofolate reductase (DHFR) moiety extended at the C terminus by a sequence containing the hemagglutinin-derived ha epitope; the Ub moiety (more specifically, the $\mathrm{Ub}^{\mathrm{R48}}$ moiety bearing the Lys → Arg alteration at position 48); a 40-residue, *E. coli* Lac repressor-derived sequence, termed e^K [extension (e) containing lysines (K)] and shown below in single-letter abbreviations for amino acids; a variable residue X between Ub and e^K; the *E. coli* βgal moiety lacking the first 24 residues of wild-type β-Gal. The lightning arrow indicates the site of *in vivo* cleavage by DUBs.[29]

N-Degron and Reporter Proteins

A change in the physiological state of a cell that is preceded or followed by the induction or repression of specific genes can be monitored through the use of promoter fusions to a variety of protein reporters, such as, for example, βgal, β-glucuronidase, luciferase, and green fluorescent protein (GFP). A long-lived reporter is useful for detecting the induction of genes, but is less suitable for monitoring either a rapid repression or a temporal pattern that involves an up- and downregulation of a gene of interest. A sufficiently short-lived reporter is required in such settings. The metabolically unstable *X*–βgal proteins of the initial N-end rule study[1] (Fig. 2) were the first such reporters. Over the last decade, other protein reporters, including those described above, were metabolically destabilized by extending them with either a portable N-degron or a "nonremovable" Ub moiety.[45–47] The latter is targeted by a distinct Ub-dependent proteolytic pathway called the UFD pathway (Ub/fusion/degradation).[1,48] These metabolically unstable proteins, expressed as Ub fusions, should be particularly useful in settings where the concentration of the reporter must reflect a recent level of gene activity. Portable N-degrons were also used to destabilize specific protein antigens, thereby enhancing the presentation of their peptides to the immune system.[49,50]

N-Degron and Conditional Mutants

A frequent problem with conditional phenotypes is their leakiness, i. e., unacceptably high residual activity of either a temperature-sensitive (*ts*) protein at nonpermissive temperature or a gene of interest in the "off" state of its promoter. Another problem is "phenotypic lag," which often occurs between the imposition of nonpermissive conditions and the emergence of a relevant null phenotype. Phenotypic lag tends to be longer with proteins that are required in catalytic rather than stoichiometric amounts.

In one application of Ub fusions and the N-end rule pathway to the problem of phenotypic lag, a constitutive N-degron (produced as a Ub fusion) was linked to a protein expressed from an inducible promoter.[51] This method is constrained by the necessity of using a heterologous promoter and by the constitutively short half-life of a target protein, whose levels may therefore be suboptimal under permissive conditions. An alternative approach is to link the N-degron to a normally long-lived protein in a strain in which the N-end rule pathway can be induced or repressed. Such strains have been constructed with *S. cerevisiae*,[52,53] but can also be designed in

---

[45] C. K. Worley, R. Ling, and J. Callis, *Plant Mol. Biol.* **37**, 337 (1998).

[46] H. Deichsel, S. Friedel, A. Detterbeck, C. Coyne, U. Hamker, and H. K. MacWilliams, *Dev. Genes Evol.* **209**, 63 (1999).

[47] I. Paz, J.-R. Meunier, and M. Choder, *Gene* **236**, 33 (1999).

[48] E. S. Johnson, P. C. Ma, I. M. Ota, and A. Varshavsky, *J. Biol. Chem.* **270**, 17442 (1995).

[49] A. Townsend, J. Bastin, K. Gould, G. Brownlee, M. Andrew, B. Coupar, D. Boyle, S. Chan, and G. Smith, *J. Exp. Med.* **168**, 1211 (1988).

[50] T. Tobery and R. F. Siliciano, *J. Immunol.* **162**, 639 (1999).

[51] E. C. Park, D. Finley, and J. W. Szostak, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1249 (1992).

[52] Z. Moqtaderi, Y. Bai, D. Poon, P. A. Weil, and K. Struhl, *Nature (London)* **383**, 188 (1996).

[53] M. Ghislain, R. J. Dohmen, F. Levy, and A. Varshavsky, *EMBO J.* **15**, 4884 (1996).

other species, including mammalian cells. The metabolic stabilities, and hence also the levels of N-degron-bearing proteins, in a cell with an inducible N-end rule pathway are either normal or low, depending on whether Ubr1p, the recognition (E3) component of the N-end rule pathway, is absent or present.[52,53] These conditional mutants can be constructed with any cytosolic or nuclear protein whose function tolerates an N-terminal extension.

Yet another design is a portable N-degron that is inactive at a low (permissive) temperature but becomes active at a high (nonpermissive) temperature. Such an N-degron was constructed, using the Ub fusion technique, from a specific *ts* allele of the 20-kDa mouse dihydrofolate reductase (DHFR) bearing the N-terminal arginine, a strongly destabilizing residue.[54] Linking this DHFR-based, heat-inducible N-degron to proteins of interest yielded a new class of *ts* mutants, called *td* (temperature-activated degron). The *td* method does not require an often unsuccessful search for a *ts* mutation in a gene of interest. If the corresponding protein can tolerate N-terminal extensions, the corresponding *td* fusion is functionally unperturbed at permissive temperature. In contrast, low activity of a *ts* protein at permissive temperature is a frequent problem with conventional *ts* mutants. The *td* method eliminates or reduces the phenotypic lag, because the activation of N-degron results in rapid disappearance of a *td* protein. Another advantage of the *td* technique is the possibility of employing two sets of conditions: a *td* protein-expressing strain at permissive versus nonpermissive temperature or, alternatively, the same strain versus a congenic strain lacking the N-end rule pathway, with both strains at nonpermissive temperature.[54] This powerful internal control, provided in the *td* technique by two alternative sets of permissive/nonpermissive conditions, is unavailable with conventional *ts* mutants. Since 1994, a few laboratories described successful uses of the *td* method to construct *ts* alleles of specific proteins (e.g., Refs. 55 and 56). A recent modification of the *td* technique combines the galactose-inducible overexulsion of Ubr1p and the temperature-sensitive (*td*) N-degron.[56a]

N-Degron and Conditional Toxins

A major limitation of the current pharmacological strategies stems from the absence of drugs that are specific for two or more independent molecular targets. For the reasons discussed in detail elsewhere,[57,58] it is desirable to have a therapeutic agent that possesses a multitarget, combinatorial selectivity, which requires the presence of two or more predetermined targets in a cell and simultaneously the absence of one or more targets for the drug to exert its effect. Note that simply combining two or more "conventional" drugs against different targets in a multidrug regimen would not yield the multitarget selectivity, because the two drugs together would perturb not only cells containing both targets but also cells containing either one of the targets.

---

[54] R. J. Dohmen, P. Wu, and A. Varshavsky, *Science* **263**, 1273 (1994).

[55] G. Caponigro and R. Parker, *Genes Dev.* **9**, 2421 (1995).

[56] J. Wolf, M. Nicks, S. Deitz, E. van Tuinen, and A. Franzusoff, *Biochem. Biophys. Res. Commun.* **243**, 191 (1998).

[56a] K. Labib, J. A. Tercero, and J. F. X. Diffley, *Science* **288**, 1643 (2000).

A strategy for designing protein-based reagents that are sensitive to the presence or absence of more than one target at the same time was proposed in 1995.[57] A key feature of these reagents is their ability to utilize codominance, the property characteristic of many signals in proteins, including degrons and nuclear localization signals (NLSs). Codominance, in this context, refers to the ability of two or more signals in the same molecule to function independently and not to interfere with each other. The critical property of a degron-based multitarget reagent is that its intrinsic toxicity is the same in all cells, whereas its half-life (and, consequently, its steady state level and overall toxicity) in a cell depends on the protein composition of the cell, specifically on the presence of "target" proteins that have been chosen to define the profile of a cell to be eliminated.[57] A related but different design involves a toxic protein made short-lived (and therefore relatively nontoxic) by the presence of a degradation signal such as an N-degron. (The latter is produced by the Ub fusion technique.) If a cleavage site for a specific viral processing protease is placed between the toxic moiety of the fusion and the N-degron, the fusion would be cleaved in virus-infected cells but not in uninfected cells. As a result, the toxic moiety of the fusion would become long-lived (and therefore more toxic) only in virus-infected cells.[59] The codominance concept and the ideas about protein-size multitarget reagents have been extended to small (<1-kDa) multitarget compounds.[58]

Overproduction of Proteins as Ubiquitin Fusions

A major application of the Ub fusion technique is its use to augment the yields of recombinant proteins.[36–39] This approach increases the yield of short peptides as well.[40,41,60] The yield-enhancing effect of Ub was observed not only with eukaryotic cells (where the Ub moiety is present in a nascent fusion but not in its mature counterpart) but also in prokaryotes, which lack the Ub system, including DUBs, and therefore retain the Ub moiety in a translated fusion.[35–37] (*Escherichia coli* transformed with a plasmid expressing the *S. cerevisiae* DUB Ubp1p acquires the ability to deubiquitylate Ub fusions.[61])

---

[57] A. Varshavsky, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3663 (1995).

[58] A. Varshavsky, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2094 (1998).

[59] A. Varshavsky, *Cold Spring Harbor Symp. Quant. Biol.* **60**, 461 (1996).

[60] T. H. LaBean, S. A. Kauffman, and T. R. Butt, *Mol. Divers.* **1**, 29 (1995).

The yield-enhancing effect of Ub stems at least in part from rapid folding of the nascent Ub moiety, whose presence at the N terminus of an emerging polypeptide chain may thereby partially protect the still unfolded chain from attacks by proteolytic pathways of the cytosol (most of these pathways are a part of the Ub system). The remarkably strong increases in protein yield even in eukaryotic cells, where the Ub moiety of the fusion is retained transiently (for it is rapidly removed by DUBs), suggest that this protection by Ub is particularly critical during translation, when an emerging, partially unfolded polypeptide chain may present degrons that are buried in the folded version of the same polypeptide. The chaperone role of Ub in this setting reflects one of its physiological functions. Specifically, the experiments with natural Ub fusions containing ribosomal proteins have shown that the transient presence of Ub in front of a ribosomal protein moiety is required for the efficient incorporation of that moiety into the nascent ribosomes,[12] most likely because of the transient protection effect described above.

The Ub-mediated increase in total yield is often accompanied by an even greater increase in the solubility of overexpressed protein. In this regard, the effect of Ub is analogous to that of several other proteins, such as thioredoxin[62] and maltose-binding protein (MBP).[63] When these moieties are cotranslationally linked to a protein of interest, they often increase its yield and solubility. A model of the underlying mechanism suggested for MBP[63] may also be relevant to the effect of Ub moiety. Specifically, a partially unfolded nascent protein is presumed to weakly interact with the nearby (upstream) MBP moiety, thereby transiently precluding intermolecular self-interactions that could result in irreversible aggregation before the protein has had the time to attain its mature conformation.[63]

The first engineered Ub fusions utilized pUB23-X, a family of high copy plasmids that expressed Ub–*X*–βgal proteins containing different junctional residues (*X*) in *S. cerevisiae* from a galactose-inducible, glucose-repressible promoter.[1,42] Subsequent designs facilitated the construction of ORFs encoding Ub–*X*–polypeptide fusions by introducing a *Sac*II (*Sst*II) site within the codons for the last three residues of the Ub moiety.[39] In this cloning scheme, an ORF of interest is amplified by polymerase chain reaction (PCR) and a primer in which the 5′ extension encodes the last three residues of Ub. Another cloning route employs double-stranded oligonucleotides with *Sac*II cohesive ends that are used to join the DNA fragments.[39] The expression of a resulting Ub–X–polypeptide fusion in a eukaryotic cell (or in a prokaryotic cell that contains the *S. cerevisiae* Ubp1p DUB) yields an X–polypeptide bearing a predetermined N-terminal residue X (Figs. 1 and 2).

---

[61] J. W. Tobias, T. E. Shrader, G. Rocap, and A. Varshavsky, *Science* **254**, 1374 (1991).

[62] E. R. LaVallie, E. A. DiBlasio, S. Kovacic, K. L. Grant, P. F. Schendel, and J. M. McCoy, *BioTechnology* **11**, 187 (1993).

[63] R. B. Kapust and D. S. Waugh, *Protein Sci.* **8**, 1668 (1999).

In their natural milieu, proteins of biotechnological or pharmacological interest are often products of the secretory pathway, and therefore are cleaved by signal peptidase on their entrance into the endoplasmic reticulum (ER). This cleavage frequently yields destabilizing residues at the N-termini of these proteins. When the same proteins are overexpressed in the cytosol of a heterologous bacterial or eukaryotic host, their N-terminal methionine tends to be retained, because MAPs cannot cleave off N-terminal methionine if it is followed by a destabilizing residue (see above). It is in these, quite frequent, cases that the expression of a protein as a Ub–X–protein fusion attains two aims at once: producing a protein of interest bearing the desired N-terminal residue (Fig. 2) and also, quite often, increasing the yield of the protein, in comparison with an otherwise identical expression of the Ub-lacking protein.[35]

There are numerous examples of Ub-mediated increases in the yield and solubility of overexpressed proteins. For instance, a conventional heterologous expression of the *Streptomyces* tyrosinase in *E. coli* yielded inactive enzyme, whereas expression of tyrosinase as a Ub fusion resulted in an abundant and active enzyme.[64] Another example of the use of Ub fusions in *E. coli* was an abundant expression of the soluble human collagenase catalytic domain. In contrast, the expression of the same protein in the absence of N-terminal Ub moiety resulted in low yield and insoluble product.[65] A 60-fold increase in the yield of the human pi class glutathione transferase GSTP1 was observed on the addition of a Ub-coding sequence to the *GSTP1* ORF.[39] A strong increase in protein yield in *E. coli* was reported with a combination of the T7 RNA polymerase promoter system and Ub fusions.[66] Several other examples of the Ub fusion approach to protein overexpression[38,67–69] are described in an earlier review by Baker.[35] More recently, Hondred and colleagues applied the Ub fusion technique to augment protein expression in transgenic plants.[70]

---

[64] K. Han, J. Hong, H. C. Lim, C. H. Kim, Y. Park, and J. M. Cho, *Ann. N.Y. Acad. Sci.* **721**, 30 (1994).

[65] M. R. Gehring, B. Condon, S. A. Margosiak, and C. C. Kan, *J. Biol. Chem.* **270**, 22507 (1995).

[66] M. H. Koken, H. H. Odijk, M. Van Duin, M. Fornerod, and J. H. Hoeijmakers, *Biochem. Biophys. Res. Commun.* **195**, 643 (1993).

Ubiquitin-Assisted Dissection of Protein Translocation across Membranes

A 1994 method called UTA (ubiquitin translocation assay) employs Ub as a kinetic probe in the context of signal sequence-bearing Ub fusions.[8] After emerging from ribosomes in the cytosol, a protein may remain in the cytosol, or may be transferred to compartments separated from the cytosolic space by membranes. With a few exceptions, noncytosolic proteins begin journeys to their respective compartments by crossing membranes that enclose intracellular organelles such as the ER and mitochondria in eukaryotes or the periplasmic space in bacteria. Amino acid sequences that enable a protein to cross the membrane of a compartment are often located at the protein's N terminus. These "signal" sequences[71] are targeted by translocation pathways specific for each compartment. The translocation of a protein across a compartment membrane can start before the synthesis of the protein is completed, resulting in docking of the still translating ribosome at the transmembrane channel. The UTA technique takes advantage of the rapid (cotranslational) cleavage of a Ub fusion to examine temporal aspects of protein transport across the ER membrane in living cells.[8] Specifically, if a Ub fusion that has been engineered to bear an N-terminal signal sequence (SS) upstream of the Ub moiety is cleaved in the cytosol by DUBs, the fusion's reporter moiety would fail to be translocated into the ER. Conversely, if a nascent SS mediates the docking of a translating ribosome at the transmembrane channel rapidly enough, or if the fusion Ub moiety is located sufficiently far downstream of the SS, then by the time the Ub moiety emerges from the ribosome the latter is already docked, and the nascent Ub moiety enters the ER before it can fold and/or be targeted by DUBs. Thus, the cleavage at the Ub moiety of an SS-bearing Ub fusion in the cytosol can serve as an *in vivo* kinetic marker and a tool for analyzing targeting in protein translocation.[8] The temporal sensitivity of the UTA technique stems from rapid folding of the nascent Ub moiety

---

[67] E. A. Sabin, C. T. Lee-Ng, J. R. Shuster, and P. J. Barr, *Bio Technology* **7**, 705 (1989).

[68] E. Rian, R. Jemtland, O. K. Olstad, J. O. Gordeladze, and K. M. Gautvik, *Eur. J. Biochem.* **213**, 641 (1993).

[69] M. Coggan, R. Baker, K. Miloszewski, G. Woodfield, and P. Board, *Blood* **9**, 2455 (1995).

[70] D. Hondred, J. M. Walker, D. E. Mathews, and R. D. Vierstra, *Plant Physiol.* **119**, 713 (1999).

[71] G. Blobel, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1496 (1980).

that precludes its translocation and makes it a substrate of DUBs in the cytosol shortly after the emergence of the fusion Ub moiety from the ribosome.

Split-Ubiquitin Sensor for Detection of Protein–Protein Interactions

Another Ub-based method, termed the split-Ub sensor or USPS (Ub/split/protein/sensor), makes it possible to detect and monitor a protein–protein interaction as a function of time, at the natural sites of this interaction in a living cell.[72] These capabilities of the split-Ub technique distinguish it from the two-hybrid assay.[73] The design of a split-Ub sensor is based on the following observations: when a C-terminal fragment of the 76-residue Ub (C_{ub}) was expressed as a fusion to a reporter protein, the fusion was cleaved by DUBs only if an N-terminal fragment of Ub (N_{ub}) was also expressed in the same cell. This reconstitution of native Ub from its fragments, detectable by the *in vivo* cleavage assay, was not observed with a mutationally altered N_{ub}. However, if C_{ub} and the altered N_{ub} were each linked to polypeptides that interact *in vivo*, the cleavage of the fusion containing C_{ub} was restored, yielding a generally applicable assay for kinetic and equilibrium aspects of the *in vivo* protein interactions.[72]

Enhancement of Ub reconstitution by interacting polypeptides linked to fragments of Ub stems from a local increase in concentration of one Ub fragment in the vicinity of the other. This in turn increases the probability that the two Ub fragments coalesce to form a quasinative Ub moiety, whose (at least) transient formation results in the irreversible cleavage of the fusion by DUBs. This cleavage can be detected readily, and can be followed as a function of time or at steady state.[72,74] Unlike the two-hybrid method, which is based on the apposition of two structurally independent protein domains whose folding and functions do not require direct interactions between the domains, the split-Ub assay involves reconstituting the conformation of a small, single-domain protein. Applications of the split-Ub sensor have shown that this assay is capable of detecting transient *in vivo* interactions such as the binding of a signal sequence of a translocated protein to Sec62p, a component of the ER channel.[74] Different reporter readouts and selection-based screens have been devised for the split-Ub assay, making it possible to use this method for identifying the *in vivo* ligands of a protein of interest, similar to the main application of the two-hybrid assay.[75,76] Split-protein sensors analogous to split-Ub but employing other proteins, such as DHFR, have been developed as well.[77,78]

---

[72] N. Johnsson and A. Varshavsky, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10340 (1994).

[73] S. Fields and O. Song, *Nature (London)* **340**, 245 (1989).

[74] M. Dünnwald, A. Varshavsky, and N. Johnsson, *Mol. Biol. Cell* **10**, 329 (1999).

UPR Technique

Direct measurements of the *in vivo* degradation of intracellular proteins require a pulse–chase assay. It involves the labeling of nascent proteins for a short time with a radioactive precursor ("pulse"), the termination of labeling through the removal of radiolabel and/or the addition of a translation inhibitor, and the analysis of a labeled protein of interest at various times afterward ("chase"), using immunoprecipitation and sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), or analogous techniques. Its advantage of being direct notwithstanding, a conventional pulse–chase assay is fraught with sources of error. For example, the immunoprecipitation yields may vary from sample to sample; the volumes of samples loaded on a gel may vary as well. If the labeling for specific chase times is done with separate batches of cells (as is the case, e.g., with anchorage-dependent mammalian cell cultures), the efficiency of labeling is yet another unstable parameter of the assay. As a result, pulse–chase data tend to be semiquantitative at best, lacking the means to correct for these errors.

A robust and convenient "internal reference" strategy was described in 1996. This strategy, an extension of the original Ub fusion method, was termed the UPR (ubiquitin/protein/reference) technique.[79] UPR can compensate for several sources of data scatter in a pulse–chase assay (Fig. 3). UPR employs a linear fusion in which Ub is located between a protein of interest and a reference protein moiety (Fig. 3A). The fusion is cotranslationally cleaved by DUBs after the last residue of Ub, producing equimolar amounts of the protein of interest and the reference protein bearing the C-terminal Ub moiety. If both the reference protein and the protein of interest are immunoprecipitated in a pulse–chase assay, the relative amounts of the protein of interest can be normalized against the reference protein in the same sample.[28,29,79] The UPR technique (Fig. 3)

---

[75] I. Stagljar, C. Korostensky, N. Johnsson, and S. te Heesen, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5187 (1998).

[76] S. Wittke, N. Lewke, S. Müller, and N. Johnsson, *Mol. Biol. Cell* **10**, 2519 (1999).

[77] I. Remy and S. W. Michnick, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5394 (1999).

[78] J. N. Pelletier, F. X. Campbell-Valois, and S. W. Michnick, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12141 (1998).

[79] F. Lévy, N. Johnsson, T. Rumenapf, and A. Varshavsky, *Proc Natl. Acad. Sci. U.S.A.* **93**, 4907 (1996).

can thus compensate for the scatter of immunoprecipitation yields, sample volumes, and other sources of sample-to-sample variation. The increased accuracy afforded by UPR underscored the insufficiency of the current "half-life" terminology, because the *in vivo* degradation of many proteins deviates from first-order kinetics. For a discussion of this problem and the terminology for describing nonexponential decay, see Refs. 29 and 79.

Ubiquitin Sandwich Technique

Nascent polypeptides emerging from the ribosome may, in the process of folding, present degradation signals similar to those recognized by the Ub system in misfolded or otherwise damaged proteins. It has been a long-standing question whether a significant fraction of nascent polypeptides is cotranslationally degraded. Determining whether nascent polypeptides are actually degraded *in vivo* has been difficult because at any given time the nascent chains of a particular protein species are of different sizes, and therefore would not form a band on electrophoresis in a conventional pulse–chase assay. The Ub sandwich technique[9] makes it possible to detect cotranslational protein degradation by measuring the steady state ratio of two reporter proteins whose relative abundance is established cotranslationally.

Operationally, the Ub sandwich technique[9] is a three-protein version of the UPR assay.[79] A polypeptide to be examined for cotranslational degradation, termed **B**, is sandwiched between two stable reporter domains **A** and **C** in a linear fusion protein. The three polypeptides are connected via Ub moieties to create a fusion protein of the form **A**Ub–**B**Ub–**C**Ub. The independent polypeptides **A**Ub, **B**Ub, and **C**Ub that result from the cotranslational cleavage of **A**Ub–**B**Ub–**C**Ub by DUBs are called modules. The DUB-mediated cleavage establishes a kinetic competition between two mutually exclusive events during the synthesis of the **A**Ub–**B**Ub–**C**Ub fusion: cotranslational UBP cleavage at the **B**Ub–**C**Ub junction to release the long-lived **C**Ub module or, alternatively, cotranslational degradation of the entire **B**Ub–**C**Ub nascent chain by the 26S proteasome. In the latter case, the processivity of proteasome-mediated degradation results in the destruction of the Ub moiety between **B** and **C** before it can be recognized by UBPs. The resulting drop in levels of the **C**Ub module relative to levels of **A**Ub, referred to as the C/A ratio, reflects the cotranslational degradation of domain **B**. This measurement provides a minimal estimate of the total amount of cotranslational degradation, because nonprocessive cotranslational degradation events that do not extend into the **C** domain are not detected. The Ub sandwich method was used to demonstrate that more than

50% of nascent protein molecules bearing an N-degron can be degraded cotranslationally in *S. cerevisiae*, never reaching their mature size before their destruction by processive proteolysis.[9]

If cotranslational protein degradation by the Ub system is found to be extensive for at least some wild-type proteins (surveys of natural proteins remain to be carried out by this new technique), it could be accounted for as an evolutionary trade-off between the necessity of identifying and destroying degron-bearing mature proteins and the mechanistic difficulty of distinguishing between posttranslationally and cotranslationally presented degrons. Cotranslational protein degradation may also represent a previously unrecognized form of protein quality control, which destroys nascent chains that fail to fold correctly. These and other questions about physiological aspects of the cotranslational protein degradation can now be addressed directly in living cells through the Ub sandwich technique.[9]

Concluding Remarks

The Ub fusion technique is made possible by the ability of DUBs to cleave a Ub fusion *in vivo* or *in vitro* after the last residue of Ub irrespective of the flanking sequence context. Since its development, the Ub fusion technique has given rise to a number of applications whose common feature is utilization of the rapid and highly specific cleavage of a Ub-containing fusion by DUBs. Among these applications is the UPR technique, which increases the accuracy of pulse–chase and analogous measurements. I hope that the use of UPR will spread, supplanting the conventional, far less accurate pulse–chase protocols that lack a reference protein. The Ub sandwich technique, a descendant of UPR, has made it possible to determine the extent of cotranslational protein degradation *in vivo* for any protein of interest. One important feature of the Ub moiety is its ability, as a part of linear fusions, to increase the yields and solubility of overexpressed proteins or short peptides in either eukaryotic or bacterial hosts. In yet another class of Ub-based applications, the demonstrated coalescence of peptide-size Ub fragments into a quasinative Ub fold has yielded the split-Ub sensor for detecting protein interactions *in vivo*. Ub fusions continue to be useful in a remarkable variety of ways.

Acknowledgments

I am most grateful to the former and current members of my laboratory, whose work made possible some of the advances described in this review. I thank Daniel Finley (Harvard Medical School) and Rohan Baker (Australian National University) for their comments on the manuscript. Our studies are supported by grants from the National Institutes of Health (GM31530 and DK39520).

The EMBO Journal Vol.18 No.23 pp.6832–6844, 1999

The E2–E3 interaction in the N-end rule pathway: the RING-H2 finger of E3 is required for the synthesis of multiubiquitin chain

Youming Xie and Alexander Varshavsky[1]

Division of Biology, 147-75, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

[1]Corresponding author
e-mail: avarsh@its.caltech.edu

We dissected physical and functional interactions between the ubiquitin-conjugating (E2) enzyme Ubc2p and Ubr1p, the E3 component of the N-end rule pathway in *Saccharomyces cerevisiae*. The binding of the 20 kDa Ubc2p by the 225 kDa Ubr1p is shown to be mediated largely by the basic residue-rich (BRR) region of Ubr1p. However, mutations of the BRR domain that strongly decrease the interaction between Ubr1p and Ubc2p do not prevent the degradation of N-end rule substrates. In contrast, this degradation is completely dependent on the RING-H2 finger of Ubr1p adjacent to the BRR domain. Specifically, the first cysteine of RING-H2 is required for the ubiquitylation activity of the Ubr1p–Ubc2p complex, although this cysteine plays no detectable role in either the binding of N-end rule substrates by Ubr1p or the physical affinity between Ubr1p and Ubc2p. These results defined the topography of the Ubc2p–Ubr1p interaction and revealed the essential function of the RING-H2 finger, a domain that is present in many otherwise dissimilar E3 proteins of the ubiquitin system.

Keywords: E2/E3/N-end rule/proteasome/RING finger/ ubiquitin

Introduction

Ubiquitin (Ub) is a 76 residue protein whose covalent conjugation to other proteins, usually in the form of a multi-Ub chain, marks these proteins for processive degradation by the 26S proteasome, an ATP-dependent multisubunit protease (Hochstrasser, 1996; Varshavsky, 1997; Baumeister *et al.*, 1998; Hershko and Ciechanover, 1998; Scheffner *et al.*, 1998). Through either constitutive or conditional degradation of many intracellular proteins, the Ub-dependent proteolytic pathways regulate a multitude of biological processes, including the cell cycle, cell differentiation, apoptosis, DNA transcription, replication and repair, signal transduction, functions of the nervous system and stress responses, including the immune response (Hicke, 1997; Varshavsky, 1997; Peters *et al.*, 1998).

The conjugation of Ub to other proteins involves the formation of a thioester between the C-terminus of Ub and a specific cysteine of the Ub-activating (E1) enzyme. The Ub moiety of E1~Ub thioester is thereafter transesterified to a specific cysteine in one of several Ub-conjugating (E2) enzymes. The Ub moiety of E2~Ub thioester is conjugated, via the isopeptide bond, to the ε-amino group of either a substrate's Lys residue or a Lys residue of another Ub moiety, the latter reaction resulting in a substrate-linked multi-Ub chain (Chau *et al.*, 1989; Pickart, 1997; Scheffner *et al.*, 1998). Most E2 enzymes function in complexes with proteins called E3s. The functions of E3s include the initial recognition of degradation signals (degrons) in the substrate proteins, with different E3s recognizing different classes of degrons. At least some E3s, specifically those containing the HECT domain, accept the Ub moiety from the associated E2~Ub thioester, forming an E3~Ub thioester and acting as a proximal donor of the Ub moiety to substrates selected by the E3 (Scheffner *et al.*, 1995; Nuber and Scheffner, 1999; Wang *et al.*, 1999).

One pathway of the Ub system is the N-end rule pathway (Bachmair *et al.*, 1986; Varshavsky, 1996). Among the targets of this pathway are proteins bearing destabilizing N-terminal residues. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues, in that they function through their conversion, by the *NTA1*-encoded N-terminal amidase, into the secondary destabilizing N-terminal residues Asp and Glu. The destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *ATE1*-encoded Arg-tRNA–protein transferase, to Arg, one of the primary destabilizing residues (Balzi *et al.*, 1990; Baker and Varshavsky, 1995). In mammals, the deamidation step is bifurcated, in that two distinct amidases specific, respectively, for N-terminal Asn or Gln, mediate the activity of tertiary destabilizing residues (Stewart *et al.*, 1995; Grigoryev *et al.*, 1996). The primary destabilizing N-terminal residues are bound directly by the *UBR1*-encoded E3 (N-recognin), the recognition component of the N-end rule pathway. In *S.cerevisiae*, Ubr1p is a 225 kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues Phe, Leu, Trp, Tyr, Ile, Arg, Lys and His (Bartel *et al.*, 1990). Ubr1p contains at least three substrate-binding sites. The type 1 site is specific for basic N-terminal residues Arg, Lys and His. The type 2 site is specific for the hydrophobic residues Phe, Leu, Trp, Tyr and Ile (Reiss *et al.*, 1988; Gonda *et al.*, 1989; Baker and Varshavsky, 1991). Ubr1p also targets short-lived proteins such as Cup9p and Gpa1p, which lack destabilizing N-terminal residues, and are recognized by Ubr1p through its third substrate-binding site, the exact location and specificity of which remain to be determined (Madura and Varshavsky, 1994; Byrd *et al.*, 1998). Similar but distinct versions of the N-end rule pathway are present in all organisms examined, from bacteria to mammals. For a summary of the currently known functions of this pathway, see Kwon *et al.* (1999a,b).

The roles of E3 proteins in the E2, E3-dependent ubiquitylation of substrates are not understood in detail.

© European Molecular Biology Organization

It has been shown that the *UBR1*-encoded N-recognin, the E3 of the N-end rule pathway, functions in a complex with Ubc2p (Rad6p) (Madura *et al.*, 1993; Watkins *et al.*, 1993), one of 13 E2 enzymes in *S.cerevisiae* (Hochstrasser, 1996). The polyacidic C-terminal tail of Ubc2p was found to be essential for the binding of Ubc2p to Ubr1p, but the Ubc2p-binding site of Ubr1p remained unknown. The mechanism of cooperation between E3 (Ubr1p) and E2 (Ubc2p) in forming a substrate-linked multi-Ub chain is unknown for either this or other Ub-dependent pathways.

We report two main results. (i) The Ubr1p–Ubc2p interaction is mediated by the basic residue-rich (BRR) region of Ubr1p and the polyacidic tail of Ubc2p. Interestingly, this high-affinity interaction is not essential for the activity of the N-end rule pathway. (ii) The RING-H2 domain of Ubr1p, adjacent to the BRR region, is strictly required for ubiquitylation of N-end rule substrates, even though this domain is not required for the binding of Ubr1p to either Ubc2p or N-end rule substrates. Because the RING-H2 and similar RING variants are a characteristic feature of many otherwise dissimilar E3 proteins, the ubiquitylation-enabling function of the Ubr1p RING-H2 identified in the present work is likely to be a general property of E3 components in the Ub system.

Terminology

Ubiquitin whose C-terminal (Gly76) carboxyl group is covalently linked to another compound is called the *ubiquityl* moiety, the derivative terms being *ubiquitylation* and *ubiquitylated*. The term Ub refers to both free ubiquitin and the ubiquityl moiety. This nomenclature, which is also recommended by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology, brings Ub-related terms in line with the standard chemical terminology (Varshavsky, 1997).

Results

Dissection of the Ubr1p–Ubc2p interaction

To search for Ubc2p-interacting proteins, we employed the two-hybrid technique (Fields and Song, 1989; James *et al.*, 1996), using the 20 kDa Ubc2p as bait. A plasmid thus identified contained a fragment of the *S.cerevisiae* open reading frame (ORF) YLR024C, which was previously termed *UBR2* (Hochstrasser, 1996; Kwon *et al.*, 1998), because the predicted sequence of the 216 kDa Ubr2p was 22% identical and 46% similar to that of the 225 kDa, *UBR1*-encoded N-recognin, the E3 of the N-end rule pathway (Bartel *et al.*, 1990). Ubr2p is not essential for cell viability under standard growth conditions and does not target N-end rule substrates, but does compete with Ubr1p for the binding to Ubc2p (H.Rao, Y.Xie and A.Varshavsky, unpublished data).

The cloned Ubc2p-interacting region of Ubr2p was a 287 residue fragment (positions 1134–1420). Sequence comparisons revealed two distinct motifs conserved between Ubr1p and Ubr2p in this region (Figure 1A). One was a basic residue-rich region, termed BRR, and the other a RING-H2 finger (Saurin *et al.*, 1996; Kwon *et al.*, 1998). The RING-H2 motif is a distinct variant of the previously defined Cys/His-rich RING motif, which is thought to be involved in protein–protein interactions (Borden and Freemont, 1996; Inouye *et al.*, 1997). Two-hybrid analyses indicated that Ubc2p interacted with the Ubr2p-homologous region of Ubr1p as well (positions 1081–1367) (Figure 1C and D). In addition, two-hybrid assays with a series of Ubr1p fragments fused to the Gal4p transcriptional activation domain demonstrated that the 287 residue region was the only detectable Ubc2p-binding site in Ubr1p (data not shown). We used site-directed mutagenesis to dissect the Ubc2p–Ubr1p interaction in this region of Ubr1p. Two sets of mutants were constructed (Figure 1B). The mutants of one set, termed MB (mutation in BRR), bore a missense mutation(s) in the BRR domain; the mutants of another set, termed MR (mutation in RING-H2) bore either Cys→Ser or His→Ala mutations in the RING-H2 finger. These derivatives of the 287 residue region of Ubr1p were tested for their binding to Ubc2p using a version of the two-hybrid assay that utilized two reporter genes, *ADE2* and *HIS3* (James *et al.*, 1996). This design increased the assay's affinity range, because a strong two-hybrid interaction was required to confer on the tester cells the Ade$^+$ phenotype, whereas a weak two-hybrid interaction sufficed to make the same tester cells His$^+$ (James *et al.*, 1996).

The replacement of a single basic residue at various positions in the BRR domain of Ubr1p with uncharged residues (e.g. the alleles MB10, MB11 and MB17) significantly weakened the interaction of Ubc2p with the 287 residue fragment of Ubr1p, as indicated by the Ade$^-$ His$^+$ phenotype of the tester strain, in comparison with the Ade$^+$ His$^+$ phenotype observed with the fragment's wild-type (wt) version (Figure 1C). When two basic residues of the BRR domain were converted to uncharged residues (MB4 allele), the resulting tester strain was still His$^+$, but grew more slowly on histidine-lacking medium than did the tester strain bearing single-substitution Ubr1p mutants (Figure 1C; data not shown), suggesting a further weakening of the Ubc2p–Ubr1p interactions. With triple-substitution mutants in the BRR domain of Ubr1p (MB5 and MB6), the affinity of Ubc2p for the 287 residue region of Ubr1p decreased to levels undetectable with the two-hybrid assay (Figure 1C).

In contrast to these findings with the BRR domain, none of the four single Cys→Ser substitutions in the adjacent RING-H2 finger of Ubr1p impaired the interaction of Ubc2p with the 287 residue region of Ubr1p (Figure 1C, mutants MR1–MR3 and MR18). A His→Ala mutation in the RING-H2 finger was also without effect (Figure 1C, mutant MR12). Moreover, even the simultaneous Cys→Ser alterations at two positions of the RING-H2 finger failed to affect the Ade$^+$ His$^+$ phenotype of the tester strain (Figure 1C, mutants MR7 and MR8). The 287 residue region of Ubr1p was then subcloned into two fusions, which contained, respectively, the BRR domain alone (positions 1081–1220) and the RING-H2 finger alone (positions 1177–1367). The RING-H2 finger alone did not exhibit affinity to Ubc2p (Figure 1D). In contrast, the binding of Ubc2p to the BRR domain alone was readily detectable, but this interaction was weakened in comparison with the interaction between Ubc2p and the 287 residue region of Ubr1p containing both BRR and RING-H2 (Figure 1D).

Taken together, these results indicated that the positively charged BRR domain is the main region mediating the binding of Ubr1p to Ubc2p, through largely electrostatic

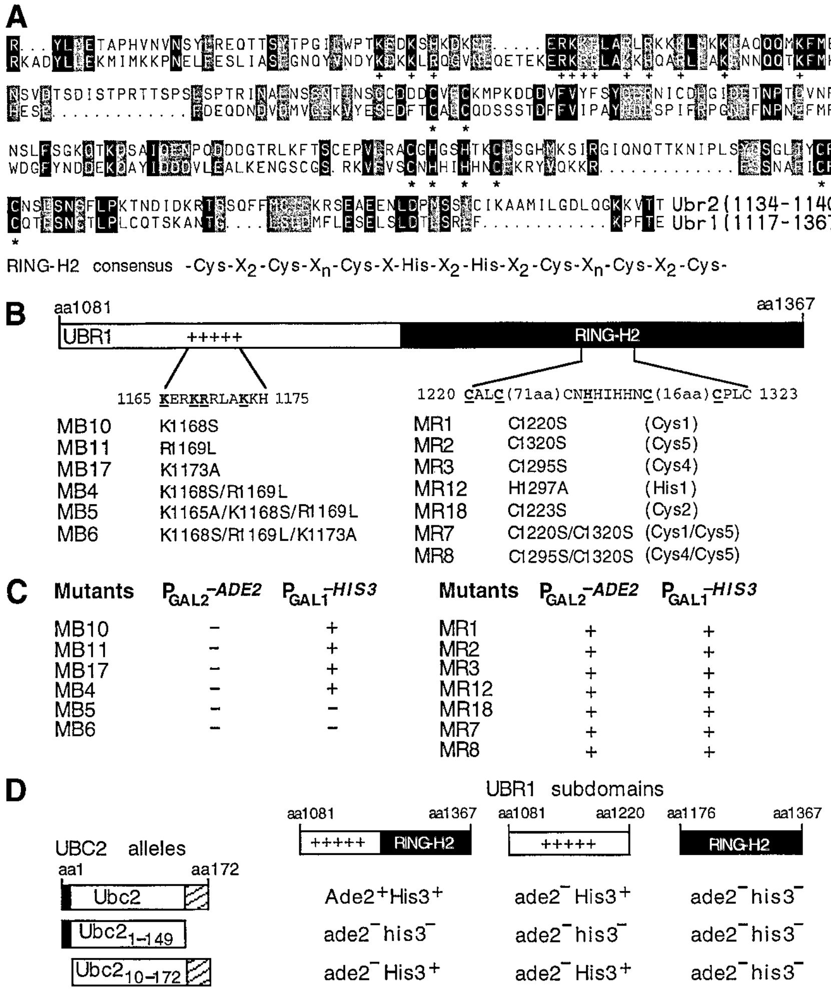

A
Ubr2 (1134-1140)
Ubr1 (1117-1367)
RING-H2 consensus -Cys-X2-Cys-Xn-Cys-X-His-X2-His-X2-Cys-Xn-Cys-X2-Cys-
B
aa1081
aa1367
UBR1
+++++
RING-H2
1165 KERKRRLAKKH 1175
1220 CALC(71aa)CNHHIHHNC(16aa)CPLC 1323
MB10 K1168S
MB11 R1169L
MB17 K1173A
MB4 K1168S/R1169L
MB5 K1165A/K1168S/R1169L
MB6 K1168S/R1169L/K1173A
MR1 C1220S (Cys1)
MR2 C1320S (Cys5)
MR3 C1295S (Cys4)
MR12 H1297A (His1)
MR18 C1223S (Cys2)
MR7 C1220S/C1320S (Cys1/Cys5)
MR8 C1295S/C1320S (Cys4/Cys5)
C
Mutants P GAL2-ADE2 P GAL1-HIS3
MB10 - +
MB11 - +
MB17 - +
MB4 - +
MB5 - -
MB6 - -
Mutants P GAL2-ADE2 P GAL1-HIS3
MR1 + +
MR2 + +
MR3 + +
MR12 + +
MR18 + +
MR7 + +
MR8 + +
D
UBR1 subdomains
aa1081 aa1367 aa1081 aa1220 aa1176 aa1367
+++++ RING-H2
+++++
RING-H2
UBC2 alleles
aa1 aa172
Ubc2
Ubc2 1-149
Ubc2 10-172
Ade2+His3+ ade2− His3+ ade2− his3−
ade2− his3− ade2− his3− ade2− his3−
ade2− His3+ ade2− His3+ ade2− his3−

Fig. 1. Ubc2p binds to the BRR domain of Ubr1p. (**A**) Sequence comparison of the Ubc2p-binding domains of *S.cerevisiae* Ubr1p and Ubr2p. Identical residues are highlighted in black; similar residues are shaded. The conserved and similar basic residues are marked by '+'. The Cys and His residues of the RING-H2 motif are denoted by asterisks. (**B**) Schematic representation of mutations in the BRR and RING-H2 domains of Ubr1p. The 287 residue region of the 1950 residue Ubr1p (positions 1081–1367) was divided into two regions containing the BRR domain and the RING-H2 finger. Mutations of basic residues, termed the MB mutations, in the BRR domain were at positions 1165–1175 and are underlined in bold. Mutations of either Cys or His residues in the RING-H2 finger, termed the MR mutations, were at positions 1220–1323 and are underlined in bold. The positions of mutated Cys and His residues within RING-H2 are listed in parentheses. (**C**) Summary of the two-reporter (*HIS3* and *ADE2*) two-hybrid results with the MB and MR mutants of Ubr1p versus wt Ubc2p. See the text and Materials and methods for details. (**D**) wt Ubc2p and its two truncated alleles, Ubc2p$_{1-149}$ and Ubc2p$_{10-172}$, were assayed in the two-reporter/two-hybrid system for their interaction with either the 287 residue Ubr1p fragment, or its BRR domain alone, or its RING-H2 domain alone. The 9 residue N-terminal region and the polyacidic C-terminal tail of Ubc2p are denoted by the black and hatched boxes, respectively.

interactions of the BRR region with the polyacidic C-terminal tail of Ubc2p (see below). These findings also indicated that the presence of the RING-H2 finger near the BRR domain was not essential for the binding of Ubc2p but was required (in an undefined way; but see below) for the high-affinity Ubr1p–Ubc2p interaction (Figure 1B–D). In addition, either single or double Cys→Ser substitutions in the RING-H2 finger did not appear to result in a gross conformational alteration of the BRR/RING-H2 region of Ubr1p, as indicated by the undiminished affinity of mutants in this region for Ubc2p (Figure 1C); further evidence for this is described below.

Previous work has shown that both the N-terminal 9 residue region and the 23 residue polyacidic C-terminal tail of Ubc2p are required for the binding of Ubc2p to Ubr1p (Madura *et al.*, 1993; Watkins *et al.*, 1993). To determine the relative contributions of these Ubc2p regions to the interaction between Ubc2p and the BRR domain of Ubr1p, we constructed two hybrid fusions containing two truncated alleles of *UBC2* which encoded, respectively, Ubc2p$_{10-172}$ (lacking residues 1–9) and Ubc2p$_{1-149}$ (lacking the polyacidic tail). These derivatives of Ubc2p were tested in the two-hybrid assay for binding to the entire 287 residue region of Ubr1p, or to the BRR domain alone, or to the RING-H2 finger alone. The deletion of the polyacidic C-terminal tail of Ubc2p abolished its interaction with the BRR domain of Ubr1p, whereas the deletion of the first 9 residues of Ubc2p weakened but did not abolish this interaction (Figure 1D). Interestingly, the latter interaction, weakened but readily detectable, was independent of the presence of the RING-H2 finger (Figure 1D). Thus, the bulk of the affinity of Ubc2p for the 287 residue fragment of Ubr1p stems from electrostatic interactions between the polyacidic C-terminal tail of Ubc2p and the positively charged, ~140 residue BRR domain of the 1950 residue Ubr1p. (The size of a

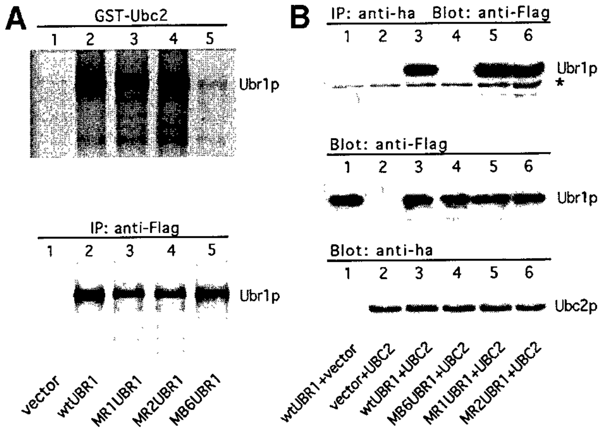


Fig. 2. The BRR domain of Ubr1p is essential for high-affinity binding to Ubc2p. (**A**) GST pull-down assays with wt Ubr1p and its mutants. FLAG-Ubr1p (denoted as wt UBR1) (lane 2), or FLAG-Ubr1p^{MR1} (lane 3) or FLAG-Ubr1p^{MR2} (lane 4), or FLAG-Ubr1p^{MB6} (lane 5) were overexpressed in AVY301 (*ubr1Δ ubc2Δ*) cells (see Materials and methods). Cells were labeled with [^{35}S]methionine for 30 min at 30°C, and the extracts were assayed using glutathione–agarose beads bound with GST–Ubc2p (top panel) or applied to immunoprecipitation with anti-FLAG monoclonal antibody (lower panel). (**B**) *In vivo* association of Ubc2p with wt Ubr1p, Ubr1p^{MR1} and Ubr1p^{MR2}, but not with Ubr1p^{MB6}. Cell extracts from transformants expressing FLAG-Ubr1p alone (lane 1) or the N-terminally HA-tagged Ubc2p (HA-Ubc2p) alone (lane 2), or co-expressing HA-Ubc2p with either FLAG-Ubr1p (lane 3) or FLAG-Ubr1p^{MB6} (lane 4), FLAG-Ubr1p^{MR1} (lane 5) or FLAG-Ubr1p^{MR2} (lane 6) were incubated with anti-HA antibody (see Materials and methods). Co-precipitated wt Ubr1p and its mutant derivatives were resolved by SDS–PAGE, followed by immunoblotting with anti-FLAG antibody (top panel). The asterisk marks a cross-reactive band. The expression levels of Ubr1p and Ubc2p in these transformants were examined by immunoblotting with, respectively, anti-FLAG (middle panel) and anti-HA antibodies (bottom panel). SDS–PAGE was carried out in 8% gels.

contiguous region within the BRR domain that is required for the binding of Ubc2p is likely to be significantly less than 140 residues.)

To determine whether basic residues of the BRR domain and the RING-H2 finger were essential for the binding of the full-length Ubr1p to Ubc2p, we first employed a pull-down assay with a fusion of Ubc2p to glutathione *S*-transferase (GST). Specifically, wt FLAG-Ubr1p (tagged at the N-terminus with the FLAG epitope), or FLAG-Ubr1p^{MB6} (a triple-substitution mutant in the BRR domain; Figure 1B), or FLAG-Ubr1p^{MR1} or FLAG-Ubr1p^{MR2} (a single Cys→Ser substitution at Cys1 and Cys5, respectively; Figure 1B) was overexpressed in a *ubr1Δ ubc2Δ S.cerevisiae* strain and labeled with [^{35}S]methionine. The N-terminal FLAG epitope did not impair the function of Ubr1p (A.Webster and A.Varshavsky, unpublished data). The extracts from labeled, Ubr1p-overexpressing cells (and the extract from cells carrying vector alone) were incubated with glutathione–agarose beads loaded with GST–Ubc2p, and the amounts of ^{35}S-labeled FLAG-Ubr1p and Ubr1p mutants recovered from the beads were determined by SDS–PAGE (Figure 2A, top panel). In parallel, the same extracts were immunoprecipitated with a monoclonal anti-FLAG antibody, to verify equality of the initial inputs of FLAG-Ubr1p and its mutants (Figure 2A, lower panel). In agreement with the results of the two-hybrid assays (Figure 1B and C), the binding of FLAG-Ubr1p^{MB6} to Ubc2p was barely detectable (Figure 2A, top panel, lane 5), whereas the binding of either FLAG-Ubr1p^{MR1} or FLAG-Ubr1p^{MR2} to Ubc2p was comparable to that of wt FLAG-Ubr1p (Figure 2A, top panel, lanes 2–4).

We also examined the *in vivo* association of wt Ubr1p and its mutants with Ubc2p, using co-immunoprecipitation–immunoblot assays. Either wt FLAG-Ubr1p, FLAG-Ubr1p^{MB6}, FLAG-Ubr1p^{MR1} or FLAG-Ubr1p^{MR2} was co-overexpressed in a *ubr1Δ ubc2Δ S.cerevisiae* with Ubc2p bearing the N-terminal hemagglutinin (HA) epitope. Controls included congenic cells expressing FLAG-Ubr1p alone or Ubc2p alone. Proteins were immunoprecipitated from cell extracts with anti-HA antibody, followed by SDS–PAGE and immunoblotting with anti-FLAG antibody (Figure 2B, top panel). The expression levels of Ubr1p and Ubc2p in these transformants were monitored by immunoblotting the initial extracts with either anti-FLAG antibody (Figure 2B, middle panel) or anti-HA antibody (Figure 2B, bottom panel). In agreement with the results of two-hybrid and pull-down assays, the *in vivo* association of FLAG-Ubr1p^{MB6} with Ubc2p was undetectable (lane 4), whereas both FLAG-Ubr1p^{MR1} and FLAG-Ubr1p^{MR2} bound Ubc2p as efficiently as wt FLAG-Ubr1p (lanes 3, 5 and 6). Thus, the BRR domain of Ubr1p is essential for the binding of Ubr1p to Ubc2p, whereas the adjacent RING-H2 finger of Ubr1p is not directly involved (Figures 1B, C and 2).

High affinity between Ubr1p and Ubc2p is not essential for degradation of N-end rule substrates

We asked whether the MB-type point mutations in the BRR domain of Ubr1p that virtually abolished the binding of Ubr1p to wt Ubc2p (Figures 1B, C and 2) affected the activity of the N-end rule pathway. wt Ubr1p and the BRR-domain mutants Ubr1p^{MB4} and Ubr1p^{MB6} (Figure 1C and D) were expressed in *ubr1Δ* cells from a low-copy plasmid and the native P_{UBR1} promoter. The *ubr1Δ* cells also expressed an N-end rule substrate, either Arg-βgal or Leu-βgal, which are short-lived in *UBR1* cells ($t_{1/2}$ of ~2 and ~3 min, respectively) (Varshavsky, 1996). Arg-βgal and Leu-βgal were expressed as Ub–Arg-βgal and Ub–Leu-βgal, which were co-translationally deubiquitylated *in vivo* (Varshavsky, 1996). Previous work has also shown that the enzymatic activity of an N-end rule substrate such as X-βgal is a sensitive measure of the substrate's metabolic stability *in vivo* (Madura *et al.*, 1993; Kwon *et al.*, 1999a).

Both Arg-βgal and Leu-βgal were long-lived *in vivo* in the absence of Ubr1p ($t_{1/2}$ >20 h) ('vector' bars in Figure 3A; data not shown). As expected, Arg-βgal and Leu-βgal were degraded in the presence of wt Ubr1p (Figure 3A). Surprisingly, the steady-state levels of Arg-βgal and Leu-βgal in the transformants with Ubr1p^{MB4} or Ubr1p^{MB6} were similar to (or only slightly higher than) the levels of these proteins in transformants expressing wt Ubr1p (Figure 3A), indicating that high-affinity binding of Ubr1p to Ubc2p is not essential for the activity of the N-end rule pathway. Control experiments with *ubc2Δ* cells confirmed that the Ubr1p^{MB6}-dependent degradation of N-end rule substrates was Ubc2p-dependent as well (data not shown).

Previous work with the tailless Ubc2p$_{1-149}$ mutant,

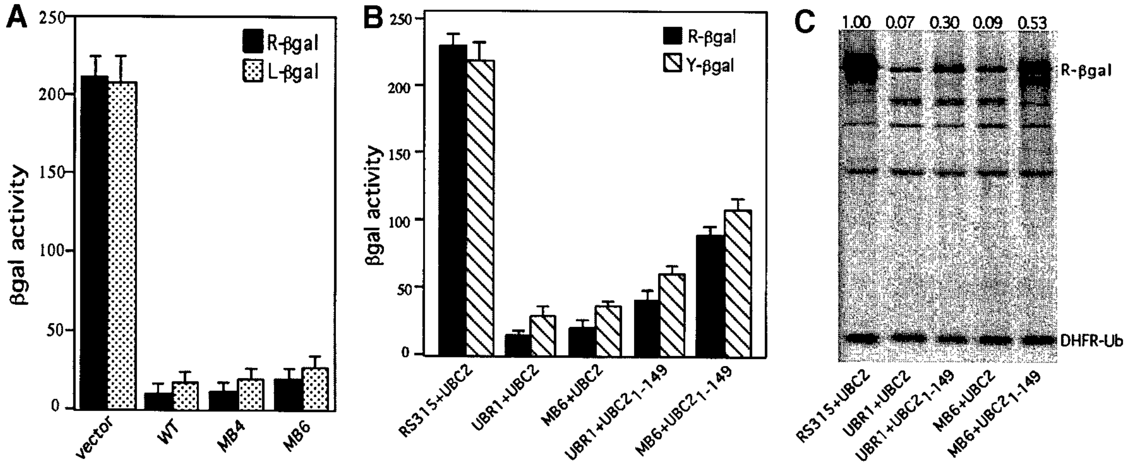


Fig. 3. High-affinity interaction between Ubr1p and Ubc2p is not essential for the activity of the N-end rule pathway. (**A**) The activity of Ubr1p is weakly affected by MB mutations. All *UBR1* alleles were expressed from the P_{UBR1} promoter in a low-copy vector. Constructs expressing wt Ubr1p, Ubr1p^{MB4} or Ubr1p^{MB6}, or the empty vector were transformed into JD55 (*ubr1*Δ) cells expressing Arg-βgal or Leu-βgal. Measurements of βgal activity were used to assay the degradation of Arg-βgal and Leu-βgal. The values shown are the means of duplicate measurements of three independent transformants. Standard deviations are indicated above the bars. (B and C) High-affinity association of Ubr1p and Ubc2p is not essential for activity of the N-end rule pathway. βgal activity assays (**B**) and immunoprecipitations with anti-βgal antibody (**C**) were used to compare the degradation of model substrates, either Arg-βgal (B and C) or Tyr-βgal (B). All *UBR1 and UBC2* alleles were expressed from the native P_{UBR1} and P_{UBC2} promoters, respectively, in a low-copy vector. Different combinations of *UBR1* and *UBC2* alleles were transformed into *ubr1*Δ *ubc2*Δ cells expressing either Arg-βgal, Tyr-βgal (B) or DHFRha–Ub^{R48}–Arg-βgal (C) (see the text and Materials and methods). The measured ratio of Arg-βgal to DHFR–Ub (the reference protein) in the absence of Ubr1p [(C), lane 1] was set at 1.00 and used to normalize the other ratios.

which did not bind to wt Ubr1p, led to the conclusion that Ubc2p$_{1-149}$ did not support the degradation of N-end rule substrates (Madura *et al.*, 1993), and implied that high-affinity binding of Ubc2p to Ubr1p was strictly essential for the pathway's function. However, another study reported a diminished but significant activity Ubc2p$_{1-149}$ (Watkins *et al.*, 1993). The latter conclusion was consistent with our findings about the BRR-region mutants of Ubr1p (Figure 3A). To address this discrepancy, we compared the activity of the N-end rule pathway in *ubr1*Δ *ubc2*Δ cells expressing different combinations of either wt Ubr1p or Ubr1p^{MB6} with either wt Ubc2p or the tailless Ubc2p (Ubc2p$_{1-149}$). *UBR1*, *UBC2* and their mutant alleles were expressed from their native promoters and a low-copy plasmid. The relative rates of degradation of Arg-βgal and Tyr-βgal ($t_{1/2}$ of ~2 and ~10 min, respectively) (Bachmair *et al.*, 1986; Varshavsky, 1996) were assayed by measuring the activity of βgal (Figure 3B). With wt Ubc2p, the activity of the N-end rule pathway did not change significantly between the wt Ubr1p and Ubr1p^{MB6} alleles (Figure 3B). Furthermore, in the presence of wt Ubr1p, the pathway's activity decreased significantly but not strongly upon the replacement of wt Ubc2p with its tailless derivative Ubc2p$_{1-149}$ (Figure 3B), in agreement with the finding of Watkins *et al.* (1993) but in contrast to the result of Madura *et al.* (1993). Interestingly, the activity of the N-end rule pathway was further decreased, but still not abolished, by combining Ubr1p^{MB6} and Ubc2p$_{1-149}$ (Figure 3B).

To measure the relative metabolic stability of Arg-βgal in a different way, we utilized a recently developed UPR (Ub/protein/reference) technique, which provides a reference protein (Lévy *et al.*, 1996, 1999). This method employs a linear fusion such as DHFRha–Ub^{R48}–X-βgal, in which Ub^{R48}, containing Arg instead of Lys at position 48, is placed between a protein of interest (X-βgal) and a reference protein moiety such as DHFRha (mouse dihydrofolate reductase bearing the HA epitope tag) (Lévy *et al.*, 1996). DHFRha–Ub^{R48}–X-βgal is co-translationally cleaved by Ub-specific deubiquitylating enzymes (DUBs) (Wilkinson and Hochstrasser, 1998) after the last residue of Ub^{R48}, producing equimolar amounts of X-βgal and DHFRha–Ub^{R48}. This way, the relative amounts of X-βgal can be normalized against the (co-immunoprecipitated) DHFRha–Ub^{R48} reference protein in the same sample. The UPR technique can therefore compensate for the scatter of immunoprecipitation yields, sample volumes, and other sources of sample-to-sample variation (Lévy *et al.*, 1996). *ubr1*Δ *ubc2*Δ cells expressing DHFRha–Ub^{R48}–Arg-βgal, and wt Ubr1p or Ubr1p^{MB6} with wt Ubc2p or Ubc2p$_{1-149}$ were labeled with [^{35}S]methionine for 30 min at 30°C. Cell extracts were immunoprecipitated with both anti-βgal and anti-HA antibodies, followed by SDS–PAGE and the determination, using a PhosphorImager, of the ratio of Arg-βgal (test protein) to DHFRha–Ub^{R48} (reference protein) (Figure 3C). The ratio determined for *UBR1*-lacking cells was arbitrarily set at 1.00, and was used to normalize the same ratios from other transformants. The results (Figure 3C) were in agreement with those obtained by measuring the enzymatic activity of βgal (Figure 3A and B). Specifically, in the presence of wt Ubc2 and Ubr1p^{MB6}, ~90% of Arg-βgal was degraded under the assay's conditions, similar to the result with wt Ubc2p and wt Ubr1p. In cells expressing wt Ubr1p and Ubc2p$_{1-149}$, ~70% of Arg-βgal was degraded. With Ubr1p^{MB6} and Ubc2p$_{1-149}$ the degradation of Arg-βgal was decreased further, to ~50%, but still not abolished (Figure 3C).

Taken together, these results (Figure 3) indicated that high-affinity interactions between the E3 Ubr1p and the E2 Ubc2p are essential for normal levels of the pathway's activity. These data also showed that even much weaker interactions between mutant variants of Ubr1p and Ubc2p, the interactions that could not be detected by either two-hybrid or pull-down assays (Figures 1 and 2), were still sufficient to sustain a large fraction of activity of the N-end rule pathway (Figure 3). We conclude that a transient interaction between Ubr1p and Ubc2p is functionally suboptimal but still sufficient for the multiubiquitylation of a substrate and for the substrate's delivery to the 26S proteasome.

The RING-H2 finger of Ubr1p is essential for degradation of N-end rule substrates

Given the proximity of the 104 residue RING-H2 finger to the BRR region of Ubr1p that mediates its binding to Ubc2p (Figure 1), we determined the effects of missense mutations in RING-H2 (Figure 1B) on the ability of Ubr1p to target N-end rule substrates for ubiquitylation and degradation. Since the structure of the folded RING-H2 domain is unknown, we constructed single missense mutants throughout this domain (Figure 1B). The sequences RING-H2^{MR1}, RING-H2^{MR18}, RING-H2^{MR3} and RING-H2^{MR2} contained single Cys→Ser substitutions at the Cys1, Cys2, Cys4 and Cys5 positions of RING-H2, respectively (Figure 1A and B). RING-H2^{MR12} contained a His→Ala substitution at the His1 position of RING-H2 (Figure 1A and B). These RING-H2 variants were introduced into wt Ubr1. The resulting mutant proteins, Ubr1p^{MR1}, Ubr1p^{MR2}, Ubr1p^{MR3}, Ubr1p^{MR12} and Ubr1p^{MR18}, were expressed in *ubr1*Δ cells from a low-copy plasmid, the native P_{UBR1} promoter, and in the presence of either Arg-βgal or Leu-βgal, as described for the BRR-domain mutants of Ubr1p (Figure 3).

In striking contrast to the negligible effect of mutations in the RING-H2 finger of Ubr1p on the binding of Ubr1p to Ubc2p (Figures 1C, D and 2), all of the tested RING-H2 mutations except MR12 had a major inhibitory effect on the degradation of N-end rule substrates (Figure 4). Specifically, Ubr1p^{MR1}, which bore a Cys→Ser substitution at Cys1 of the RING-H2 finger, was completely inactive as an E3 (N-recognin), in that the steady-state levels of Arg-βgal and Leu-βgal in cells expressing Ubr1p^{MR1} were indistinguishable from the levels of these substrates in *ubr1*Δ cells, which lacked the N-end rule pathway (Figure 4). Other mutations of the RING-H2 finger impaired the N-end rule pathway strongly but incompletely, whereas the MR12 mutation (His→Ala at His1 of RING-H2) had a relatively weak effect (Figure 4).

Previous work has shown that Ubr1p is a rate-limiting component of the N-end rule pathway, in that overexpression of Ubr1p accelerated the degradation of N-end rule substrates (Bartel *et al.*, 1990). We introduced the apparently inactive Ubr1p^{MR1} and the partially active Ubr1p^{MR2} mutations into wt Ubr1p bearing the N-terminal FLAG epitope and expressed the resulting proteins from the P_{ADH1} promoter on a high-copy plasmid. Overexpression of the partially active Ubr1p^{MR2} increased the activity of the N-end rule pathway in comparison with cells expressing Ubr1p^{MR2} at wt levels (Figure 4). In contrast, the effect of overexpressing Ubr1p^{MR1} was marginal, in that only ~4% of Leu-βgal and <20% of Arg-βgal were degraded, in comparison with virtually no degradation at physiological levels of Ubr1p^{MR1} (Figure 4). Finally, the ~90 kDa *in vivo* cleavage product of βgal (denoted by an asterisk), which is characteristic of short-lived X-βgal substrates and is not produced in the absence of either Ubr1p or Ubc2p (Dohmen *et al.*, 1991), was nearly absent with Ubr1p^{MR1} (Figures 4B and 5B); this result was consistent with negligible functional activity of Ubr1p^{MR1}.

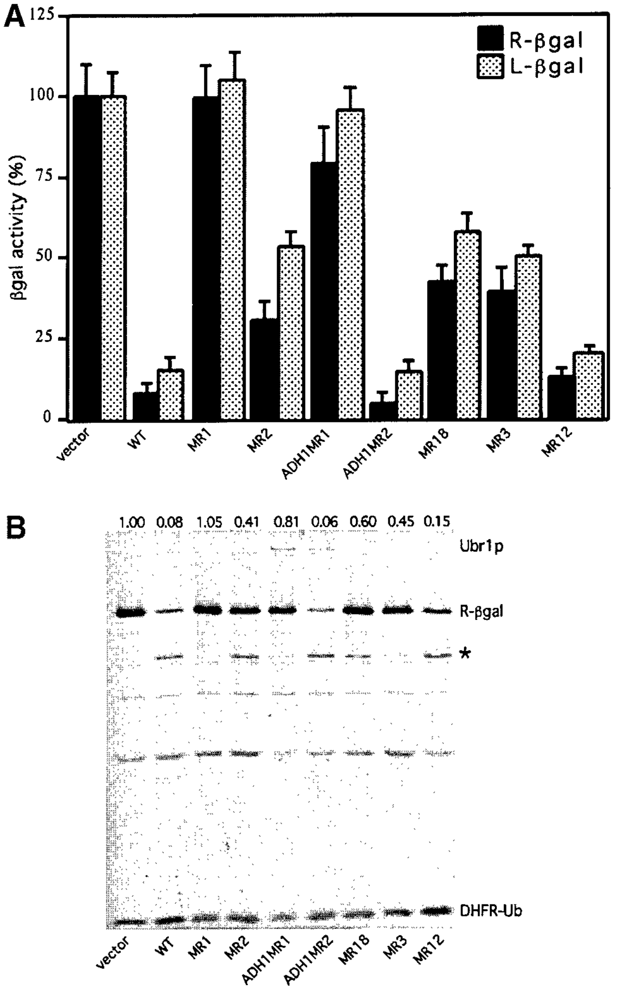


Fig. 4. Proteolysis by the N-end rule pathway requires the RING-H2 finger of Ubr1p. Methods described in the legend to Figure 3 and Materials and methods were used to assess the degradation of test substrates Arg-βgal (A and B) and Leu-βgal (A) in the presence of different *UBR1* alleles (see Figure 1B). All *UBR1* alleles were expressed from the native P_{UBR1} promoter in a low-copy vector. The Ubr1p^{MR1} (ADH1MR1) and Ubr1p^{MR2} (ADH1MR2) mutants were also expressed from the P_{ADH1} promoter in a high-copy vector. Constructs expressing wt Ubr1p and its mutants were transformed into JD55 (*ubr1*Δ) cells expressing the Arg-βgal or Leu-βgal [(**A**), βgal activity assays], or DHFRha-Ub^{R48}-Arg-βgal [(**B**), immunoprecipitation assays]. The activity of βgal in the absence of *UBR1* (vector alone) was set at 100%. The values shown are the means of duplicate measurements of three independent transformants. Standard deviations are indicated above the bars. In (B), the asterisk denotes the 90 kDa βgal cleavage product characteristic of the short-lived X-βgals (Bachmair *et al.*, 1986). Ubr1p co-immunoprecipitated with anti-βgal antibody is also indicated. The measured ratio of Arg-βgal to DHFR-Ub in the absence of Ubr1p [(B), lane vector] was set at 1.00 and used to normalize the other ratios.

Previous work has shown that the N-end rule pathway controls the import of peptides through degradation of the homeodomain protein Cup9p, which inhibits the expression of the Ptr2p peptide transporter (Byrd *et al.*, 1998). We examined whether cells expressing Ubr1p^{MR1} instead of wt Ubr1p were perturbed in their ability to import peptides by comparing the growth of *lys2*Δ *ubr1*Δ cells

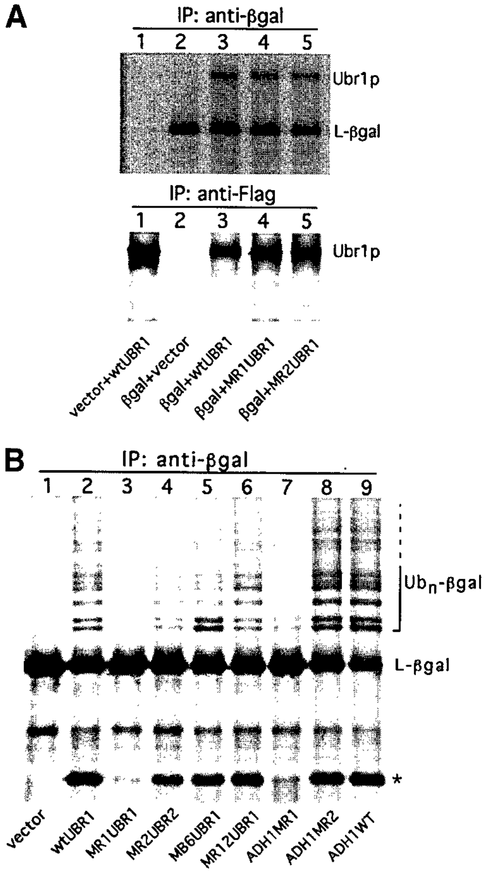


Fig. 5. The RING-H2 finger of Ubr1p is required for ubiquitylation of N-end rule substrates. (**A**) The RING-H2 finger is not required for the binding of N-end rule substrates by Ubr1p. The UPR-based fusion DHFRha–Ub^{R48}–Leu-βgal (see the text and Materials and methods) was co-expressed in AVY301 (*ubr1Δ ubc2Δ*) cells with the FLAG-tagged wt Ubr1p (lane 3), Ubr1p^{MR1} (lane 4), Ubr1p^{MR2} (lane 5) or vector alone (lane 2). The Ubr1p proteins were expressed from the P$_{ADH1}$ promoter in a high-copy vector. Lane 1, AVY301 cells expressing wt Ubr1p alone. Extracts from cells labeled with [^{35}S]methionine were immunoprecipitated with anti-βgal (top panel) or anti-FLAG antibodies (lower panel), followed by SDS–8% PAGE (see Materials and methods). (**B**) The RING-H2 finger of Ubr1p is required for ubiquitylation of N-end rule substrates. JD55 (*ubr1Δ*) cells expressing Leu-βgal and different amounts of Ubr1p or its mutants were labeled with [^{35}S]methionine for 6 min at 30°C, followed by the addition of NEM to 50 mM and preparation of extracts by heating at 100°C in the presence of SDS (see Materials and methods). The extracts were precipitated with anti-βgal antibody and fractionated by SDS–6% PAGE. Note the virtual absence of the multiubiquitylated Leu-βgal species (denoted Ub$_n$-βgal on the right) from cells expressing Ubr1p^{MR1} (lane 3). The asterisk denotes the 90 kDa βgal cleavage product characteristic of short-lived X-βgals (Bachmair *et al.*, 1986).

expressing either wt Ubr1p, Ubr1p^{MR1} or Ubr1p^{MB6} on plates containing Lys–Ala dipeptide as the sole source of lysine. Under these conditions, inactivation of the N-end rule pathway is lethal, owing to the inability of cells (auxotrophic for lysine) to import the Lys–Ala dipeptide (Byrd *et al.*, 1998). Cells expressing Ubr1p^{MR1}, which is virtually inactive as N-recognin, failed to grow on Lys–Ala dipeptide plates, whereas cells expressing either wt Ubr1p or Ubr1p^{MB6} survived (data not shown), consistent with the effects of the MR1 and MB6 mutations on the activity of the N-end rule pathway (Figures 3 and 4).

We conclude that at least Cys1 of the RING-H2 finger in Ubr1p is essential for the E3 function of Ubr1p in the N-end rule pathway, even though this cysteine is not required for the binding of Ubr1p to Ubc2p. Yet another insight from the pulse–immunoprecipitation assays with Arg-βgal was that the anti-βgal antibody co-precipitated the partially active FLAG-Ubr1p^{MR2} and the inactive FLAG-Ubr1p^{MR1} from the extracts of cells that overexpressed these Ubr1p mutants (Figure 4B). Thus, the functionally inactivating Cys1→Ser1 mutation in the RING-H2 finger of Ubr1p^{MR1} may not have impaired the binding of substrates by Ubr1p^{MR1}. This conjecture is confirmed below.

RING-H2 of Ubr1p is not required for the binding of N-end rule substrates by Ubr1p

To compare directly the affinities of Ubr1p and its derivatives for an N-end rule substrate, FLAG-Ubr1p, FLAG-Ubr1p^{MR1} or FLAG-Ubr1p^{MR2} were overexpressed in a *ubr1Δ ubc2Δ* strain, which also expressed Leu-βgal (produced from Ub–Leu-βgal). Owing to the absence of Ubc2p, Leu-βgal was long-lived in this strain, irrespective of the presence of Ubr1p (Dohmen *et al.*, 1991). Controls included congenic cells expressing FLAG-Ubr1p alone or Leu-βgal alone. Cells were labeled with [^{35}S]methionine for 30 min; proteins were immunoprecipitated from extracts with anti-βgal antibody, followed by SDS–PAGE. Anti-βgal co-immunoprecipitated Leu-βgal and comparable amounts of either wt FLAG-Ubr1p, FLAG-Ubr1p^{MR1} or FLAG-Ubr1p^{MR2} (Figure 5A, upper panel, lanes 3–5). This co-immunoprecipitation was specific for Leu-βgal, because no FLAG-Ubr1p was brought down with anti-βgal in the absence of Leu-βgal (Figure 5A, upper panel, lane 1). The expression levels of FLAG-Ubr1p, FLAG-Ubr1p^{MR1} and FLAG-Ubr1p^{MR2} were nearly identical, as could be shown by immunoprecipitating these proteins with anti-FLAG antibody (Figure 5A, lower panel). In contrast to the results with anti-βgal antibody (Figure 5A, upper panel), Leu-βgal was not co-precipitated with the FLAG-tagged Ubr1p proteins by anti-FLAG antibody (Figure 5A, lower panel). The probable explanation is that under these conditions, only free (overproduced) FLAG-Ubr1p proteins could be immunoprecipitated by anti-FLAG, because in the Leu-βgal/FLAG-Ubr1p complex Leu-βgal is bound to the type 2 substrate-binding site of Ubr1p, located in the N-terminal region of Ubr1p (A.Webster, M.Ghislain and A.Varshavsky, unpublished data). As a result, the N-terminus of Ubr1p could be shielded from recognition by the bulky βgal moiety bound to Ubr1p.

Taken together, these results (Figures 1–5) indicated that the MR1 (Cys1→Ser1) mutation in the RING-H2 finger of Ubr1p virtually eliminated the E3 function of Ubr1p, but did not impair its affinity for either N-end rule substrates or the Ubc2p Ub-conjugating enzyme.

The RING-H2 finger of Ubr1p is required for ubiquitylation of N-end rule substrates

To examine ubiquitylation of an N-end rule substrate in cells containing mutant Ubr1p proteins, FLAG-Ubr1p^{MR1}, FLAG-Ubr1p^{MR2}, FLAG-Ubr1p^{MR12}, FLAG-Ubr1p^{MB6} (Figure 1B) and the wt FLAG-Ubr1p were expressed in a *ubr1Δ UBC2* strain, which also expressed Leu-βgal (produced from Ub–Leu-βgal). To minimize the extent of the degradation of Leu-βgal during labeling, the duration of the [^{35}S]methionine pulse was reduced to 6 min. At the end of the pulse, high concentrations of *N*-ethylmaleimide (NEM), the inhibitor of DUBs (Johnsson and Varshavsky, 1994), were employed to inhibit deubiquitylation after cell lysis. ^{35}S-labeled, immunoprecipitated Leu-βgal was fractionated by SDS–PAGE (Figure 5B).

Multiubiquitylated derivatives of Leu-βgal were observed in the presence but not in the absence of wt Ubr1p, and the extent of ubiquitylation increased upon overexpression of FLAG-Ubr1p (Figure 5B, lanes 1, 2 and 9). However, virtually no ubiquitylated Leu-βgal was observed with Ubr1p^{MR1}, the degradation-inactive RING-H2 mutant of Ubr1p (Figure 5B, lane 3, compare with lanes 1 and 2). Cells overexpressing FLAG-Ubr1p^{MR1} contained trace amounts of multiubiquitylated Leu-βgal, in comparison with much higher levels of these species in cells overexpressing wt FLAG-Ubr1p (Figure 5B, lane 7, compare with lane 9). In agreement with the results of degradation tests (Figure 4), the partially degradation-defective RING-H2 mutant Ubr1p^{MR2} yielded intermediate levels of the multiubiquitylated Leu-βgal, and ubiquitylation was strongly enhanced upon overexpression of FLAG-Ubr1p^{MR2} (Figure 5B, lanes 4 and 8). Ubr1p^{MR12}, whose ability to mediate the degradation of N-end rule substrate was similar to that of wt Ubr1p (Figure 4), yielded the ubiquitylation pattern indistinguishable from that produced by wt Ubr1p (Figure 5B, lane 6). Consistent with the results of βgal activity assays (Figure 3), Ubr1p^{MB6}, which bore three mutations in the BRR domain and was severely impaired in its binding to Ubc2p (Figures 1–3), yielded nearly wt levels of multiubiquitylated Leu-βgal (Figure 5B, lane 5, compare with lane 2). Taken together, these results indicated that the RING-H2 finger of Ubr1p is specifically required for the Ubr1p, Ubc2p-dependent multiubiquitylation of N-end rule substrates, and that Cys1 of the RING-H2 finger is a functionally critical residue of this domain.

Dominant-negative effect of the ubiquitylation-inactive RING-H2 mutant of Ubr1p

The earlier gel filtration data suggested that Ubr1p functions as a monomer (Bartel, 1990). In addition, the immunoprecipitation of FLAG-Ubr1p with anti-FLAG antibody from extracts of cells that overexpressed both FLAG-Ubr1p and Ubr1p-HA did not co-immunoprecipitate Ubr1p-HA, also suggesting the monomeric configuration of Ubr1p (A.Webster and A.Varshavsky, unpublished data). We asked whether the functionally inactive Ubr1p^{MR1} could exert a dominant-negative effect on the N-end rule pathway. Either wt Ubr1p or Ubr1p^{MR1} (in addition to vector alone) was overexpressed from the P$_{ADH1}$ promoter on a high-copy plasmid in *UBR1* cells that also expressed the UPR-based fusions which produced Arg-βgal (or Leu-βgal) and equimolar amounts of DHFR–Ub^{R48}, the reference protein. Cells were labeled with [^{35}S]methionine for 30 min, followed by immunoprecipitation of Leu-βgal (Figure 6A) or Arg-βgal (Figure 6B), as well as the reference protein DHFR–Ub^{R48}. The ratios of X-βgal to DHFR–Ub^{R48} were measured, and thereafter normalized by the same ratios in congenic *ubr1*Δ cells (Figure 6A and B, lane 1). Whereas overexpresion of wt Ubr1p in *UBR1* cells accelerated the degradation of Leu-βgal and Arg-βgal, overexpression of Ubr1p^{MR1} strongly inhibited this degradation (Figure 6A and B, lanes 2 and 3, compare with lanes 4). The dominant-negative effect of overexpressed Ubr1p^{MR1} was caused at least in part by the sequestration of endogenous Ubc2p, inasmuch as this effect could be partially reversed through overexpression of Ubc2p (data not shown). The sequestration of N-end rule substrates by Ubr1p^{MR1}, which could bind to substrates indistinguishably from wt Ubr1p (Figure 5), was another likely (and independent) cause of the dominant-negative effect of Ubr1p^{MR1} (Figure 6). The possibility that overexpression of Ubr1p^{MR1} might also interfere with the downstream components of the Ub system, such as the 26S proteasome, was ruled out by the observation that the degradation of Ub–Pro-βgal was not affected by the overexpression of Ubr1p^{MR1} (data not shown). Ub–Pro-βgal is a short-lived protein degraded by the UFD (Ub/fusion/degradation) pathway, which targets the non-removable (or slowly removable) N-terminal Ub moiety of a fusion protein (Johnson *et al.*, 1995; Koegl *et al.*, 1999).

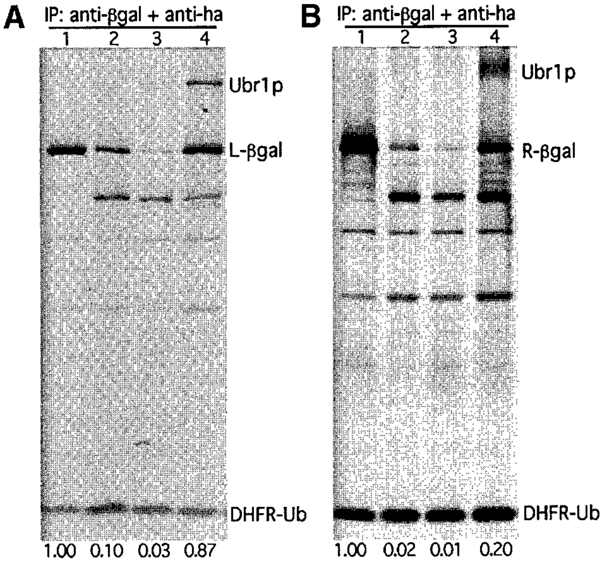


Fig. 6. Dominant-negative effect of Ubr1p^{MR1}, the RING-H2 mutant of Ubr1p. The P$_{ADH1}$ promoter-based, high-copy vector alone (lane 2), and its derivatives expressing the FLAG-tagged wt Ubr1p (lane 3) or Ubr1p^{MR1} (lane 4) were transformed into JD52 (*UBR1*) cells expressing the UPR-based fusion DHFRha–Ub^{R48}–Leu-βgal (**A**) or DHFRha–Ub^{R48}–Arg-βgal (**B**). Cells were labeled with [^{35}S]methionine for 30 min at 30°C, followed by extraction, immunoprecipitation with a mixture of anti-βgal and anti-HA antibodies, and SDS–12% PAGE, and quantitation. Lanes 1 in (A) and (B) show the results of the same tests with JD55 (*ubr1*Δ) cells in the absence of any version of Ubr1p. The measured ratio of the test protein (Leu-βgal or Arg-βgal) to the reference protein (DHFRha–Ub^{R48}) in lanes 1 was set at 1.00. The ratios determined in the other settings and normalized against the lane-1 ratios are shown below the lanes.

Discussion

We report two main results. First, we identified and dissected the E2 enzyme-binding site of Ubr1p, the E3 component of Ub-protein ligase in the N-end rule pathway of *S.cerevisiae*. This site, the BRR domain, is shown to bind the Ubc2p Ub-conjugating (E2) enzyme mainly through electrostatic interactions between the basic residues of BRR and the polyacidic C-terminal tail of Ubc2p. Although high-affinity interactions between the E3 Ubr1p and the E2 Ubc2p are essential for normal activity of the N-end rule pathway, even much weaker interactions between mutant variants of Ubr1p and Ubc2p, the interactions which could not be detected by either two-hybrid or pull-down assays, were found to be sufficient for sustaining a large fraction of the pathway's activity. Secondly, we discovered the requirement for the RING-H2 finger of Ubr1p in the ubiquitylation of substrates mediated by the Ubr1p–Ubc2p complex. A single Cys→Ser missense mutation in the RING-H2 finger (which is adjacent to the BRR domain in Ubr1p) severely impairs ubiquitylation and degradation of N-end rule substrates, but does not impair the ability of Ubr1p to bind the substrates and Ubc2p. These findings defined the physical and functional aspects of the Ubc2p–Ubr1p interaction, and revealed a specific essential function of the RING-H2 finger. This domain is present not only in the members of the UBR family of E3s (Kwon *et al.*, 1998) but also in otherwise dissimilar E3 proteins of the Ub system.

Physical and functional aspects of E2–E3 interaction

Ubr1p, the 225 kDa E3 protein, was shown to interact with the 20 kDa Ubc2p E2 enzyme mainly, if not exclusively, through the BRR domain of <15 kDa. This interaction requires the polyacidic C-terminal tail of Ubc2p and has a strong electrostatic component, because conversions of single basic residues in the BRR domain into neutral ones resulted in comparable decreases in the affinity of Ubr1p for Ubc2p. Moreover, this affinity was decreased in rough proportion to the number of mutated basic residues in BRR (Figures 1 and 2). These results are consistent with the reported dependence of the Ubr1p–Ubc2p interaction on the length of the polyacidic tail of Ubc2p. Specifically, the truncated Ubc2p derivative that lacked eight of 23 residues of the polyacidic tail retained high affinity for Ubr1p, but further shortening of the Ubc2p polyacidic tail strongly decreased the affinity (Madura *et al.*, 1993).

Previous work suggested that the 9 residue N-terminal region of Ubc2p was essential for the binding of Ubc2p to Ubr1p (Watkins *et al.*, 1993). Our current data demonstrated that deletion of the first 9 residues of Ubc2p weakened but did not eliminate its interaction, in the two-hybrid assay, with the 287 residue fragment of Ubr1p that contained the BRR domain and the RING-H2 finger (Figure 1). It is not clear how the N-terminal 9 residues of Ubc2p contribute to the interaction with Ubr1p. Inasmuch as Arg6 and Arg8 of the folded Ubc2p are likely to interact through a hydrogen bond (Worthylake *et al.*, 1998), it is possible that the effect of the 9 residue N-terminal deletion is indirect, stemming from a conformational destabilization of Ubc2p. Our finding of the 287 residue Ubr1p fragment that contains the BRR and RING-H2 domains and is sufficient for the high-affinity binding to Ubc2p (Figures 1 and 2) should facilitate the dissection of this E2–E3 interaction at the atomic level through co-crystallization of the previously solved Ubc2p (Worthylake *et al.*, 1998) and the 287 residue fragment of Ubr1p.

Function of the RING-H2 finger of Ubr1p

One of the major results of this work is the discovery of an essential and specific requirement for the RING-H2 finger of Ubr1p in the ubiquitylation of N-end rule substrates mediated by the Ubr1p–Ubc2p complex. The RING-H2 motif is a distinct variant of the previously defined Cys/His-rich RING motif. In several proteins outside the Ub system, RING-H2 has been shown to function as a site of protein–protein interactions (Borden and Freemont, 1996; Inouye *et al.*, 1997; Kamura *et al.*, 1999). In addition to the UBR family of E3 proteins (Kwon *et al.*, 1998), several other E3 components of the Ub system, including Apc11p, Roc1p (Rbx1p or Hrt1p), Roc2p and their homologs in multicellular eukaryotes, have also been found to contain RING-H2 or other variants of RING (Yu *et al.*, 1998; Zachariae *et al.*, 1998; Ohta *et al.*, 1999; Seol *et al.*, 1999; Skowyra *et al.*, 1999; Tan *et al.*, 1999).

The RING-H2 of Ubr1p is adjacent to the BRR domain, which mediates the physical interaction between Ubr1p and Ubc2p (Figures 1 and 2). A single Cys→Ser mutation at position 1 of the RING-H2 finger abolished ubiquitylation and degradation of N-end rule substrates, but did not impair the ability of the resulting mutant, Ubr1p^{MR1}, to bind both the substrates and Ubc2p (Figures 4–6). The Cys→Ser mutations at other positions of the RING-H2 finger had similar but partial effects, while the His→Ala mutation at His1 position of RING-H2 had the smallest effect among the mutants tested (Figures 4–6). Thus, even though the Ubc2p E2 enzyme can interact with the E1 enzyme to yield the Ubc2p~Ub thioester, and can also produce isopeptide bond-mediated Ub–protein conjugates in the absence of E3 (Jentsch *et al.*, 1987), the synthesis of a multi-Ub chain by the Ubc2p–Ubr1p complex requires the presence of cysteine at position 1 of the RING-H2 finger in Ubr1p.

The Ub system of a cell contains a number of distinct E2 and E3 proteins (Hochstrasser, 1996; Varshavsky, 1997; Hershko and Ciechanover, 1998). An E3 protein of a Ub-dependent proteolytic pathway was initially presumed to function as a non-catalytic, substrate-recognition component of the corresponding Ub-protein ligase, which consists of an E3 and its associated E2 enzyme. Ubr1p (N-recognin) of the present work was the first E3 that has been shown to bind proteins bearing specific degradation signals (Bachmair *et al.*, 1986; Bartel *et al.*, 1990). More recently, it was found that in addition to recognizing specific substrates, at least some E3s, specifically those that contain the HECT domain (Scheffner *et al.*, 1993), also function as enzymes. An E3 of this class can accept the Ub moiety from an associated E2~Ub thioester, initially forming the E3~Ub thioester confined to a specific (and essential) cysteine of E3. The Ub moiety of this E3~Ub thioester is thereafter conjugated, through the isopeptide bond, either directly to a substrate protein or to the Ub moiety of a growing, substrate-linked multi-Ub chain (Scheffner *et al.*, 1995). Other classes of E3 proteins, including the SCF, APC and UBR families (Hershko and Ciechanover, 1998; Kwon *et al.*, 1998; Tyers and Willems, 1999), lack recognizable HECT domains, and have not thus far been demonstrated to form thioesters with Ub.

The mutations tested in the RING-H2 finger of Ubr1p (including the Cys1 mutation which abolished the ubiquitylation activity of the Ubr1p–Ubc2p complex) did not impair the ability of Ubr1p to bind the Ubc2p enzyme and N-end rule substrates (Figures 1–5). Thus, the specific functional effect of these mutations cannot be due to a global conformational destabilization of Ubr1p. We conclude that the Ubr1p (E3) component of the N-end rule pathway not only recognizes specific degradation signals in substrate proteins, but in addition directly participates, through a mechanism that requires the RING-H2 domain, in the Ubc2p-mediated synthesis of a substrate-linked multi-Ub chain. In the first of possible models, a specific cysteine of RING-H2, perhaps the essential Cys1 residue (Figure 1), acts as an acceptor of the Ub moiety transesterified from the Ubc2p~Ub thioester, similar to the one previously described mechanism of the HECT-domain E3 called E6AP (Scheffner *et al.*, 1995; Huibregtse *et al.*, 1998). Earlier experiments with an *in vitro* ubiquitylation system containing purified *S.cerevisiae* Ubr1p and other targeting components of the N-end rule pathway suggested the Ubc2p-dependent formation of a Ubr1p~Ub thioester (V.Chau and A.Varshavsky, unpublished data). Attempts to rigorously verify the functionally relevant presence of such a thioester in the completely defined, Ubr1p-based *in vitro* system have not been successful thus far (F.Du and A.Varshavsky, unpublished data). Another possible role of the cysteine(s) of the RING-H2 finger in Ubr1p may be to modify the immediate vicinity of the active cysteine (Cys88) of the Ubr1p-associated Ubc2p in a way that facilitates the multi-Ub synthesis by Ubc2p, in the absence of Ubr1p~Ub thioester formation. Physical proximity between the Ubc2p-binding BRR domain of Ubr1p and its RING-H2 domain (Figure 1A) is consistent with this mechanism. A third model posits the existence of an unidentified factor that binds to the RING-H2 finger of Ubr1p, and is required for the efficient synthesis of multi-Ub by the Ubr1p–Ubc2p complex. Jentsch and colleagues (Koegl *et al.*, 1999) have identified a factor, termed E4, which facilitates the synthesis of a multi-Ub chain linked to a substrate of the *S.cerevisiae* UFD pathway. This factor, encoded by the *UFD2* gene, is required for the *in vivo* degradation of Ub^{V76}–Val-βgal, but is not required for the degradation of a much smaller and monomeric UFD substrate such as Ub^{V76}–Val-DHFR; Ufd2p is also not required for the degradation of βgal-based N-end rule substrates (Johnson *et al.*, 1995). Whether the function of the RING-H2 finger of Ubr1p is mediated by a finger-binding protein analogous to E4 remains to be determined.

The RING-H2 finger and its variants are present in several otherwise dissimilar classes of E3 proteins in the Ub system. For example, the mammalian E3 encoded by the *Mdm2* gene and specific for the p53 substrate contains a putative RING finger (Elenbaas *et al.*, 1996); the conserved Cys464 residue of this domain was proposed to define an active site analogous to the Ub thioester-forming site of the HECT-domain E3 enzymes (Honda *et al.*, 1997). Roc1p (Rbx1p, Hrt1p), a recently identified subunit of the cell cycle-regulating Ub–protein ligase complex called SCF, contains a variant of RING-H2 in which Cys6 is replaced by Asp (Kamura *et al.*, 1999). Mutagenesis-based analyses suggested that the RING finger of Roc1p was required for the function of the SCF complex (Kamura *et al.*, 1999; Ohta *et al.*, 1999). The sequence of Roc1p (Rbx1p, Hrt1p) is highly similar to that of Apc11p, the RING-H2-containing subunit of another cell cycle-regulating Ub–protein ligase, called APC or cyclosome (Yu *et al.*, 1998; Zachariae *et al.*, 1998). The RING-H2 finger and its variants have also been found in other (possibly also E3) components of the Ub system, including Siah-1 and Der3p (Hrd1p), which are required, respectively, for the degradation of the DCC protein (deleted in colorectal cancer) (Hu and Fearon, 1999) and of the misfolded lumenal and integral membrane proteins (Bordallo *et al.*, 1998).

Further dissection of the RING domains in these and other E3 proteins will determine whether the ubiquitylation-enabling function of RING-H2 in Ubr1p of the N-end rule pathway is a general property of the E3 components in the Ub system.

Materials and methods

Strains, media and β-galactosidase assay

The *S.cerevisiae* strains used were JD55 (MAT**a** *trp1-Δ63 ura3-52 his3-Δ200 leu2-3, 112 lys2-801 ubr1Δ::HIS3*); JD52 (MAT**a** *trp1-Δ63 ura3-52 his3-Δ200 leu2-3, 112 lys2-801*); AVY301 (a *ubc2Δ::URA3* derivative of JD55); and PJ69-4A (a gift from P.James and E.Craig) (MAT**a** *trp1-109 leu2-3, 112 ura3-52 his3-Δ200 gal4Δ gal80Δ GAL2-ADE2 LYS2::GAL1-HIS3 met::GAL7-lacZ*). Cultures were grown in rich (YPD) or in synthetic media containing standard ingredients (Sherman, 1991) and either 2% glucose (SD medium), 2% galactose (SG medium) or 2% raffinose. Yeast were transformed as described previously (Chen *et al.*, 1992). For the induction of the P_{CUP1} promoter, $CuSO_4$ was added to a final concentration of 0.1 mM (Ghislain *et al.*, 1996). The enzymatic activity of βgal in yeast extracts, from cultures at OD_{600} of 0.8–1.0, was determined as described previously (Baker and Varshavsky, 1991), using *o*-nitrophenyl-β-D-galactopyranoside (ONPG). The *Escherichia coli* strain used was DH5α.

Uses of the two-hybrid system

The *UBC2* ORF was amplified by PCR using pKM1313 (Madura and Varshavsky, 1994) as the template DNA. The two primers were designed so that the PCR products contained a *Nco*I site immediately 5′ to the start codon, and a *Sal*I site immediately 3′ to the stop codon. The *Nco*I–*Sal*I-cut PCR fragment was subcloned into *Nco*I–*Sal*I-cut PAS2 (Durfee *et al.*, 1993), yielding pAS2-UBC2, which expressed a fusion containing the DNA-binding domain of Gal4p upstream of the Ubc2p moiety. The construct was verified by nucleotide sequencing, and also by immunoblotting, in which the Gal4–Ubc2p protein was detected with an anti-Gal4 antibody (Santa Cruz Biotechnology, Santa Cruz, CA). Gal4–Ubc2p retained the ability of Ubc2p to rescue the UV-sensitivity defect of *ubc2Δ* cells. The PCR primers used were ggaaccatggagatgtccacaccagctagaag and ccttgtcgactcagtctgcttcgtcgtcgt. The two-hybrid DNA libraries and the host *S.cerevisiae* strain PJ69-4A were gifts from P.James and E.Craig (James *et al.*, 1996). PJ69-4A expressed three reporters from different inducible promoters: P_{GAL1}-*HIS3*, P_{GAL2}-*ADE2*, and P_{GAL7}-*lacZ*. pAS2-UBC2 was transformed into PJ69-4A. Trp$^+$ transformants were then transformed with the two-hybrid libraries, with selection directly on SD (–trp, –leu, –ade) plates at 30°C. Ade$^+$ colonies were then tested on SD (–trp, –leu, –his) plates. Library plasmids were rescued from the Ade$^+$ His$^+$ transformants, and re-transformed into PJ69-4A expressing either pAS2-UBC2 or the control vector pAS2. The plasmids that induced *ADE2* and *HIS3* in the presence but not in the absence of pAS2-UBC2-encoded putative Ubc2p-interacting proteins, and were analyzed by partial sequencing. One clone thus identified encoded a 287 residue fragment (position 1134–1420) of the ORF YLR024C, which was previously termed *UBR2* (Hochstrasser, 1996; Kwon *et al.*, 1998), as described in Results.

The two-hybrid system was also used to qualitatively assess the interaction of either Ubc2p or its truncated derivatives with the BRR domain and/or the RING-H2 finger of either wt Ubr1p or its derivatives containing point mutations. $UBC2_{1-149}$, encoding the C-terminally truncated Ubc2p lacking the 23 residue polyacidic tail, was constructed using PCR and the primers ggaaccatggagatgtccacaccagctagaag and gctagtcgacccaagatttctctaccgtctc. The PCR products were cut by *Sal*I and *Bam*HI, and subcloned into *Sal*I–*Bam*HI-cut pGBDUC1 (James *et al.*, 1996), yielding $pGBDUC1UBC2_{1-149}$, which expressed a fusion of $Ubc2p_{1-149}$ to the DNA binding domain of Gal4p. Similarly, $UBC2_{10-172}$, encoding Ubc2p lacking the first 9 residues, was constructed using the primers tacgggatccatgagagattttaaacgtatg and ccttgtcgactcagtctgcttcgtcgtcgt. The PCR products were digested with *Sal*I and *Bam*HI, and subcloned into *Sal*I–*Bam*HI-cut pGBDUC1 vector, yielding $pGBDUC1UBC2_{10-172}$. The *UBR1* fragment encoding residues 1081–1367, including the BRR domain and the RING-H2 finger, was produced using PCR and the primers aactggatccggtaaggactctataccagaagct and ttccgtcgacctcagtaaaaggtttgaagagcct. The PCR products were cut with *Bam*HI and *Sal*I, and subcloned into *Bam*HI–*Sal*I-cut pGADC1 (James *et al.*, 1996), yielding pGADC1UBR1$_{1081-1367}$, which encoded a fusion of Ubr1p$_{1081-1367}$ to the transcriptional activation domain of Gal4p. *UBR1*$_{1081-1220}$, encoding the Ubr1p fragment containing the BRR domain but not the RING-H2 finger, was constructed using the primers tacgggatccatgagagattttaaacgtatg and ttccgtcgacgcaggtaaaatcctccgattc, the above procedure, and yielding pGADC1UBR1$_{1081-1220}$, which expressed a fusion of Ubr1p$_{1081-1220}$ to the transcriptional activation domain of Gal4. *UBR1*$_{1177-1367}$, encoding the Ubr1p fragment containing the RING-H2 finger but not the BRR domain, was constructed using the primers aactggatccgcaaggcttctagccaag and ttccgtcgacgcaggtaaaatcctccgattc, the above procedure, and yielding pGADC1UBR1$_{1177-1367}$, which expressed a fusion of Ubr1p$_{1177-1367}$ to the transcriptional activation domain of Gal4. *UBR1* alleles containing point mutations in either the BRR domain or the RING-H2 finger were produced using the QuickChange mutagenesis kit (Stratagene, La Jolla, CA). Briefly, an ~850 bp *Eco*RI–*Bgl*II fragment from pGADC1UBR1$_{1081-1367}$ that encoded the BRR domain and the RING-H2 finger was subcloned into the *Eco*RI–*Bgl*II-cut LITMUS28 vector (New England Biolabs, Beverly, MA), yielding LITMUS28UBR1(*Eco*RI–*Bgl*II), which was subjected to site-directed mutagenesis. The resulting plasmids were cut with *Eco*RI and *Bgl*II, and the ~850 bp *Eco*RI–*Bgl*II fragments containing the intended point mutations were subcloned back into *Eco*RI–*Bgl*II-cut pGADC1UBR1$_{1081-1367}$, yielding pGADC1UBR1MB10, pGADC1UBR1MB11, pGADC1UBR1MB17, pGADC1UBR1MB4, pGADC1UBR1MB5, and pGADC1UBR1MB6; pGADC1UBR1MR1, pGADC1UBR1MR2, pGADC1UBR1MR3, pGADC1UBR1MR12, pGADC1UBR1MR18, pGADC1UBR1MR7 and pGADC1UBR1MR8. All PCR products and inserts containing point mutations were verified by DNA sequencing on both strands. These mutants belonged to either the MB or the MR groups (Figure 1B).

UPR plasmids encoding N-end rule substrates

UPR plasmids were the yeast vector-based counterparts of the plasmids originally constructed for expression in mammalian cells (Lévy *et al.*, 1996). These plasmids, termed pRβgalUPR and pLβgalUPR, expressed, respectively, DHFR–Ub–Arg-βgal and DHFR–Ub–Leu-βgal from the P$_{\mathrm{CUP1}}$ promoter. A *Bam*HI–*Sma*I insert of pBARUPR or pBALUPR was replaced with a 3.8 kb βgal-encoding insert from pUB23-M (Bachmair *et al.*, 1986), using *Bam*HI and *Sca*I. pBARUPR and pBALUPR were derived from the low-copy, pRS314-based pBAM vector (Ghislain *et al.*, 1996). They expressed, respectively, DHFR–Ub–Arg-Ura3p and DHFR–Ub–Leu-Ura3p from the P$_{\mathrm{CUP1}}$ promoter (A.Webster and A.Varshavsky, unpublished data), and were a gift from A.Webster.

Plasmids encoding Ubr1p, Ubc2p and their mutant derivatives

UBR1 and its mutant alleles were expressed either from its natural promoter in a low-copy, pRS315-derived vector or from the P$_{\mathrm{ADH1}}$ promoter in a high-copy, YEplac181-based plasmid. Briefly, pRS315 (Sikorski and Hieter, 1989) was cut with *Xba*I, blunted by Klenow Pol I, and thereafter self-ligated, yielding pRS315(–*Xba*I) lacking the *Xba*I site. pUBR1 (Bartel *et al.*, 1990) was cut with *Sal*I and *Pst*I. An ~8.1 kb insert containing the ORF and promoter of *UBR1* was subcloned into *Sal*I–*Pst*I-cut pRS315(–*Xba*I), yielding pRS315(–*Xba*I)UBR1. To produce mutant *ubr1* alleles, a ~2.9 kb *Bgl*II–*Xba*I fragment of *UBR1* from pRS315(–*Xba*I)UBR1 was subcloned into LITMUS28. The resulting LITMUS28UBR1(*Bgl*II–*Xba*I) was digested with *Hpa*I and *Sap*I, and an ~1.4 kb fragment was isolated. LITMUS28UBR1(*Bgl*II–*Xba*I) was also cut by *Sap*I, and an ~3.7 kb band was isolated. The 1.4 and 3.7 kb fragments were ligated with a set of ~600 bp *Sap*I–*Hpa*I fragments from the point mutation-containing pGADC1UBR1MB4, pGADC1UBR1MB6, pGADC1UBR1MR1, pGADC1UBR1MR2, pGADC1UBR1MR3, pGADC1UBR1MR12 or pGADC1UBR1MR18, yielding, respectively, pLITSUM28UBR1MB4, pLITMUS28UBR1MB6, pLITMUS28UBR1MR1, pLITMUS28UBR1MR2, pLITMUS28UBR1MR3, pLITMUS28UBR1MR12 and pLITMUS28UBR1MR18. A 2.9 kb *Bgl*II–*Xba*I fragment from each of these plasmids was subcloned into *Bgl*II–*Xba*I-cut pRS315(–*Xba*I)UBR1, yielding, respectively, pRS315(–*Xba*I)UBR1MB4, pRS315(–*Xba*I)UBR1MB6, pRS315 (–*Xba*I)UBR1MR1, pRS315(–*Xba*I)UBR1MR2, pRS315(–*Xba*I)UBR1MR3, pRS315(–*Xba*I)UBR1MR12 and pRS315(–*Xba*I)UBR1MR18. pNTFlagUBR1 expressed the full-length Ubr1p bearing the N-terminal FLAG epitope (MDYKDDDDK) from the P$_{\mathrm{ADH1}}$ promoter in a high-copy, LEU2-containing vector. The MB6, MR1 or MR2 mutations (Figure 1B) were introduced into pNTFlagUBR1 by replacing its 2.9 kb *Bgl*II–*Xba*I insert with its counterparts from pRS315(–*Xba*I)UBR1MB6, pRS315(–*Xba*I)UBR1MR1 or pRS315(–*Xba*I)UBR1MR2, yielding, respectively, pNTFlagUBR1MB6, pNTFlagUBR1MR1 and pNTFlagUBR1MR2.

UBC2 and *UBC2*$_{1-149}$ alleles containing the native P$_{\mathrm{UBC2}}$ promoter (~500 bp of the 5′ region) were amplified by PCR. The *Sac*I–*Sal*I-cut PCR products were subcloned into *Sac*I–*Sal*I-cut p413GAL1 (Mumberg *et al.*, 1994). Both wt Ubc2p and Ubc2p$_{1-149}$ expressed from these vectors were able to complement the UV-sensitivity phenotype of *ubc2*Δ cells (data not shown). The PCR primers for the wt *UBC2* and the *ubc2*$_{1-149}$ were, respectively, acgtgagctcccctaattgaattgttaagcg and ccttgtcgactcagtctgcttcgtcgtcgt, and acgtgagctcccctaattgaattgttaagcg and gctagtcgactcaccaagatttctctaccgtctc.

GST binding assays

The purified GST–Ubc2p fusion protein (~1 μg) was diluted into 0.5 ml of the loading buffer (1% Triton X-100, 10% glycerol, 0.5 M NaCl, 1 mM EDTA, 50 mM Tris–HCl pH 8.0), incubated with 10 μl (bed volume) of the glutathione–agarose beads (Sigma, St Louis, MO) for 1 h at 4°C, and washed three times with 1 ml of the binding buffer (10% glycerol, 50 mM NaCl, 50 mM Na–HEPES pH 7.5). Washed beads were incubated with extracts from ^{35}S-labeled cells (~5 × 10^6 CCl$_3$COOH-insoluble counts/min/extract) containing the vector alone or a plasmid expressing one of the *UBR1* alleles, including wt FLAG-Ubr1p (specifically, pNTFlagUBR1, pNTFlagUBR1MB6, pNTFlagUBR1MR1 or pNTFlagUBR1MR2) at 4°C for 1 h. The incubation buffer contained 1 mg/ml of the ovalbumin carrier (Sigma, St Louis, MO) in the binding buffer. Beads were washed three times with the binding buffer. Bound proteins were eluted with the loading buffer containing 2% SDS and fractionated by SDS–8% PAGE.

Co-immunoprecipitation–immunoblotting

Ubc2p bearing the N-terminal HA epitope was expressed from the P$_{\mathrm{GAL1}}$ promoter in a low-copy number vector p416GAL1 (Mumberg *et al.*, 1994). AVY301 cells (*ubr1*Δ *ubc2*Δ) containing a plasmid expressing HA-tagged Ubc2p or a plasmid expressing FLAG-Ubr1p, or plasmids expressing HA-tagged Ubc2p and one of the *UBR1* alleles, including pNTFlagUBR1, pNTFlagUBR1MB6, pNTFlagUBR1MR1 or pNTFlagUBR1MR2 were cultured in raffinose medium to an OD$_{600}$ of 0.8–1.0, then switched to galactose medium for induction of Ubc2p expression for 6 h. Cells were harvested and resuspended in DB buffer (50 mM NaCl, 1 mM EDTA, 50 mM Na–HEPES pH 7.5) containing 1× protease inhibitor mix (Boehringer Mannheim), and thereafter lysed by vortexing with glass beads (Bartel *et al.*, 1990). For the co-immunoprecipitation of Ubc2p with Ubr1p (Figure 2B, top panel), supernatants were incubated with anti-HA antiserum (Babco, Berkeley, CA), followed by SDS–8% PAGE and immunoblotting with a monoclonal anti-FLAG antibody (Sigma, St Louis, MO). The expression levels of Ubr1p and Ubc2p in these transformants were monitored by immunoblotting with anti-FLAG antibody (Figure 2B, middle panel) and anti-HA antibody (Figure 2B, bottom panel), respectively. SDS–8% PAGE was used.

Metabolic labeling and immunoprecipitation

Saccharomyces cerevisiae cells from 10 ml cultures (OD$_{600}$ 0.8–1.2) in SD media containing 0.1 mM CuSO$_4$ were labeled for 30 min at 30°C with 0.15 mCi of [^{35}S]methionine/cysteine (EXPRESS, New England Nuclear, Boston, MA). The same conditions were used to label FLAG-Ubr1p and its derivatives for the GST pull-down assays, except that no CuSO$_4$ was added. To assay the *in vivo* ubiquitylation of N-end rule substrates, cells were labeled for 6 min, followed by the addition of 2 M NEM in dimethylformamide to a final concentration of 50 mM (Johnsson and Varshavsky, 1994). Two methods were used to prepare cell extracts. For the GST pull-down assays, cell extracts were prepared as described above for co-immunoprecipitation immunoblots. For the experiments shown in Figures 3C, 4B, 5A and 6, labeled cells were pelleted, and resuspended in 0.8 ml of buffer A (1% Triton X-100, 0.15 M NaCl, 1 mM EDTA, 50 mM Na–HEPES pH 7.5) containing protease inhibitors, and were lysed as above. The extracts were centrifuged at 12 000 *g* for 10 min, and supernatants containing equal amounts of CCl$_3$COOH-insoluble ^{35}S were used for the GST assays described above or for immunoprecipitations. To detect Ubr1p (Figures 2A and 5A, lower panels), immunoprecipitations were carried out with a monoclonal anti-FLAG antibody conjugated to agarose beads (Kodak, Rochester, NY), followed by SDS–8% PAGE. For the co-immunoprecipitation of Leu-βgal with Ubr1p (Figure 5A, top panel), supernatants were incubated with a monoclonal anti-βgal antibody (Promega, Madison, WI), followed by SDS–8% PAGE. For measurements of X-βgal stability using the

UPR technique (Figures 3C, 4B and 6), a mixture of an anti-HA antiserum (Babco, Berkeley, CA) and the monoclonal anti-βgal antibody was incubated with ^{35}S-labeled cell extracts. Immunoprecipitated proteins were separated by SDS–12% PAGE. For detection of the *in vivo* ubiquitylation, cells labeled for 6 min were lysed using the SDS-boiling method (Hochstrasser and Varshavsky, 1990). Briefly, a pellet of NEM-treated cells (see above) was mixed with an equal volume of the 2× SDS buffer at 100°C (2% SDS, 30 mM dithiothreitol, 90 mM Na–HEPES pH 7.5) plus 50 mM NEM, and incubated at 100°C for 5 min. The extracts were diluted 10-fold with buffer A plus protease inhibitors and 50 mM NEM. CCl_3COOH-insoluble ^{35}S was determined, and samples containing equal amounts of ^{35}S were immunoprecipitated with the monoclonal anti-βgal antibody, followed by SDS–6% PAGE.

Acknowledgements

We are grateful to F.Du and A.Webster for the GST–Ubc2p fusion protein and the pNTFlagUBR1 and pUPR plasmids. We thank E.Craig and P.James for their two-hybrid constructs, and members of the Varshavsky laboratory, especially F.Du, H.Rao, H.-R.Wang, J.Sheng and G.Turner, for helpful discussions and comments on the manuscript. This work was supported by grants to A.V. from the National Institutes of Health (DK39520 and GM31530). Y.X. was supported in part by a postdoctoral fellowship from the NIH.

References

Bachmair,A., Finley,D. and Varshavsky,A. (1986) *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science*, **234**, 179–186.

Baker,R.T. and Varshavsky,A. (1991) Inhibition of the N-end rule pathway in living cells. *Proc. Natl Acad. Sci. USA*, **87**, 2374–2378.

Baker,R.T. and Varshavsky,A. (1995) Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway. *J. Biol. Chem.*, **270**, 12065–12074.

Balzi,E., Choder,M., Chen,W., Varshavsky,A. and Goffeau,A. (1990) Cloning and functional analysis of the arginyl-tRNA–protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **265**, 7464–7471.

Bartel,B. (1990) Molecular genetics of the ubiquitin system: the ubiquitin fusion proteins and proteolytic targeting mechanisms. PhD thesis. Department of Biology, MIT, Cambridge, MA.

Bartel,B., Wünning,I. and Varshavsky,A. (1990) The recognition component of the N-end rule pathway. *EMBO J.*, **9**, 3179–3189.

Baumeister,W., Walz,J., Zühl,F. and Seemüller,E. (1998) The proteasome: paradigm of a self-compartmentalizing protease. *Cell*, **92**, 367–380.

Bordallo,J., Plemper,R.K., Finger,A. and Wolf,D.H. (1998) Der3p/Hrd1p is required for endoplasmic reticulum-associated degradation of misfolded lumenal and integral membrane proteins. *Mol. Biol. Cell*, **9**, 209–222.

Borden,K.L. and Freemont,P.S. (1996) The RING finger domain: a recent example of a sequence-specific family. *Curr. Opin. Struct. Biol.*, **6**, 395–401.

Byrd,C., Turner,G.C. and Varshavsky,A. (1998) The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. *EMBO J.*, **17**, 269–277.

Chau,V., Tobias,J.W., Bachmair,A., Marriott,D., Ecker,D.J., Gonda,D.K. and Varshavsky,A. (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science*, **243**, 1576–1583.

Chen,D.-C., Yang,B.-C. and Kuo,T.-T. (1992) One-step transformation of yeast in stationary phase. *Curr. Genet.*, **21**, 83–84.

Dohmen,R.J., Madura,K., Bartel,B. and Varshavsky,A. (1991) The N-end rule is mediated by the UBC2 (RAD6) ubiquitin-conjugating enzyme. *Proc. Natl Acad. Sci. USA*, **88**, 7351–7355.

Durfee,T., Becherer,K., Chen,P.L., Yeh,S.H., Yang,Y., Kilbum,A.E., Lee,W.H. and Elledge,S.J. (1993) The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. *Genes Dev.*, **7**, 555–569.

Elenbaas,B., Dobbelstein,M., Roth,J., Shenk,T. and Levine,A.J. (1996) The MDM2 oncoprotein binds specifically to RNA through its RING finger domain. *Mol. Med.*, **2**, 439–451.

Fields,S. and Song,O. (1989) A novel genetic system to detect protein–protein interactions. *Nature*, **340**, 245–246.

Ghislain,M., Dohmen,R.J., Levy,F. and Varshavsky,A. (1996) Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in *Saccharomyces cerevisiae*. *EMBO J.*, **15**, 4884–4899.

Gonda,D.K., Bachmair,A., Wünning,I., Tobias,J.W. and Varshavsky,A. (1989) Universality and structure of the N-end rule. *J. Biol. Chem.*, **264**, 16700–16712.

Grigoryev,S., Stewart,A.E., Kwon,Y.T., Arfin,S.M., Bradshaw,R.A., Jenkins,N.A., Copeland,N.G. and Varshavsky,A. (1996) A mouse amidase specific for N-terminal asparagine. The gene, the enzyme and their function in the N-end rule pathway. *J. Biol. Chem.*, **271**, 28521–28532.

Hershko,A. and Ciechanover,A. (1998) The ubiquitin system. *Annu. Rev. Biochem.*, **76**, 425–479.

Hicke,L. (1997) Ubiquitin-dependent internalization and down-regulation of plasma membrane proteins. *FASEB J.*, **11**, 1215–1226.

Hochstrasser,M. (1996) Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.*, **30**, 405–439.

Hochstrasser,M. and Varshavsky,A. (1990) *In vivo* degradation of a transcriptional regulator: the yeast α2 repressor. *Cell*, **61**, 697–708.

Honda,R., Tanaka,H. and Yasuda,H. (1997) Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. *FEBS Lett.*, **420**, 25–27.

Hu,G. and Fearon,E.R. (1999) Siah-1 N-terminal RING domain is required for proteolysis function and C-terminal sequences regulate oligomerization and binding to target proteins. *Mol. Cell. Biol.*, **19**, 724–732.

Huibregtse,J.M., Maki,C.G. and Howley,P.M. (1998) Ubiquitination of the p53 tumor suppressor. In Peters,J.-M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell*. Plenum Press, New York, NY, pp. 323–343.

Inouye,C., Dhillon,N. and Thorner,J. (1997) Ste5 RING-H2 domain: role in Ste4 promoted oligomerization for yeast pheromone signaling. *Science*, **278**, 103–106.

James,P., Halladay,J. and Craig,E. (1996) Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics*, **144**, 1425–1436.

Jentsch,S., McGrath,J.P. and Varshavsky,A. (1987) The yeast DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme. *Nature*, **329**, 131–134.

Johnson,E.S., Ma,P.C., Ota,I.M. and Varshavsky,A. (1995) A proteolytic pathway that recognizes ubiquitin as a degradation signal. *J. Biol. Chem.*, **270**, 17442–17456.

Johnsson,N. and Varshavsky,A. (1994) Ubiquitin-assisted dissection of protein transport across cell membranes. *EMBO J.*, **13**, 2686–2698.

Kamura,K. *et al.* (1999) Rbx1, a component of the VHL tumor suppressor complex and SCF ubiquitin ligase. *Science*, **284**, 657–661.

Koegl,M., Hoppe,T., Schlenker,S., Ulrich,H.D., Mayer,T.U. and Jentsch,S. (1999) A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. *Cell*, **96**, 635–644.

Kwon,Y.T. *et al.* (1998) The mouse and human genes encoding the recognition component of the N-end rule pathway. *Proc. Natl Acad. Sci. USA*, **95**, 7898–7903.

Kwon,Y.T., Kashina,A.S. and Varshavsky,A. (1999a) Alternative splicing results in differential expression, activity and localization of the two forms of arginyl-tRNA–protein transferase, a component of the N-end rule pathway. *Mol. Cell. Biol.*, **19**, 182–193.

Kwon,Y.T., Lévy,F. and Varshavsky,A. (1999b) Bivalent inhibitor of the N-end rule pathway. *J. Biol. Chem.*, **274**, 18135–18139.

Lévy,F., Johnsson,N., Rumenapf,T. and Varshavsky,A. (1996) Using ubiquitin to follow the metabolic fate of a protein. *Proc. Natl Acad. Sci. USA*, **93**, 4907–4912.

Lévy,F., Johnston,J.A. and Varshavsky,A. (1999) Analysis of a conditional degradation signal in yeast and mammalian cells. *Eur. J. Biochem.*, **259**, 244–252.

Madura,K. and Varshavsky,A. (1994) Degradation of Gα by the N-end rule pathway. *Science*, **265**, 1454–1458.

Madura,K., Dohmen,R.J. and Varshavsky,A. (1993) N-recognin/Ubc2 interactions in the N-end rule pathway. *J. Biol. Chem.*, **268**, 12046–12054.

Mumberg,D., Muller,R. and Funk,M. (1994) Regulatable promoters of *Saccharomyces cerevisiae*—comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.*, **22**, 5767–5768.

Nuber,U. and Scheffner,M. (1999) Identification of determinants in E2 ubiquitin conjugating enzymes required for hect E3 ubiquitin–protein ligase interactions. *J. Biol. Chem.*, **274**, 7576–7582.

Ohta,T., Michel,J.J., Schottelius,A.J. and Xiong,Y. (1999) Roc1, a homolog of Apc11, represents a family of cullin partners with an associated ubiquitin ligase activity. *Mol. Cell*, **3**, 535–541.

Peters,J.-M., Harris,J.R. and Finley,D. (eds) (1998) *Ubiquitin and the Biology of the Cell.* Plenum Press, New York, NY.

Pickart,C.M. (1997) Targeting of substrates to 26S proteasome. *FASEB J.*, **11**, 1055–1066.

Reiss,Y., Kaim,D. and Hershko,A. (1988) Specificity of binding of N-terminal residues of proteins to ubiquitin–protein ligase. Use of amino acid derivatives to characterize specific binding sites. *J. Biol. Chem.*, **263**, 2693–2698.

Saurin,A.J., Borden,K.L.B., Boddy,M.N. and Freemont,P.S. (1996) Does this have a familiar RING? *Trends Biochem. Sci.*, **21**, 208–214.

Scheffner,M., Huibregtse,J.M., Vierstra,R.D. and Howley,P.M. (1993) The HPV-16 E6 and E6-AP complex functions as a ubiquitin–protein ligase in the ubiquitination of p53. *Cell*, **75**, 495–505.

Scheffner,M., Nuber,U. and Huibregtse,J.M. (1995) Protein ubiquitination involving an E1–E2–E3 enzyme ubiquitin thioester cascade. *Nature*, **373**, 81–83.

Scheffner,M., Smith,S. and Jentsch,S. (1998) The ubiquitin conjugation system. In Peters,J.-M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell.* Plenum Press, New York, NY, pp. 65–98.

Seol,J.H. *et al.* (1999) Cdc53/cullin and the essential Hrt1 RING-H2 subunit of SCF define a ubiquitin ligase module that activates the E2 enzyme Cdc34. *Genes Dev.*, **13**, 1614–1626.

Sherman,F. (1991) Getting started with yeast. *Methods Enzymol.*, **194**, 3–21.

Sikorski,R.S. and Hieter,P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *S. cerevisiae. Genetics*, **122**, 19–27.

Skowyra,D., Koepp,D.M., Kamura,T., Conrad,M.N., Conaway,R.C., Conaway,J.W., Elledge,S.J. and Harper,J.W. (1999) Reconstitution of G_1 cyclin ubiquitination with complexes containing SCFGrr1 and Rbx1. *Science*, **284**, 662–665.

Stewart,A.E., Arfin,S.M. and Bradshaw,R.A. (1995) The sequence of porcine protein NH_2-terminal asparagine amidohydrolase. A new component of the N-end rule pathway. *J. Biol. Chem.*, **270**, 25–28.

Tan,P., Fuchs,S.Y., Chen,A., Wu,K., Gomez,C., Ronai,Z. and Pan,Z.-Q. (1999) Recruitment of Roc1–Cul1 ubiquitin ligase by Skp1 and HOS to catalyze the ubiquitination of IκBα. *Mol. Cell*, **3**, 527–533.

Tyers,M. and Willems,A.R. (1999) One ring to rule a superfamily of E3 ubiquitin ligases. *Science*, **284**, 602–604.

Varshavsky,A. (1996) The N-end rule: functions, mysteries, uses. *Proc. Natl Acad. Sci. USA*, **93**, 12142–12149.

Varshavsky,A. (1997) The ubiquitin system. *Trends Biochem. Sci.*, **22**, 383–387.

Wang,G., Yang,J. and Huibregtse,J.M. (1999) Functional domains of the Rsp5 ubiquitin–protein ligase. *Mol. Cell. Biol.*, **19**, 342–352.

Watkins,J.F., Sung,P., Prakash,S. and Prakash,L. (1993) The extremely conserved amino-terminus of RAD6 ubiquitin-conjugating enzyme is essential for N-end rule-dependent protein degradation. *Genes Dev.*, **7**, 250–261.

Wilkinson,K. and Hochstrasser,M. (1998) The deubiquitinating enzymes. In Peters,J.M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell.* Plenum Press, New York, NY, pp. 99–125.

Worthylake,D.K., Prakash,S., Prakash,L. and Hill,C.P. (1998) Crystal structure of the *Saccharomyces cerevisiae* ubiquitin-conjugating enzyme Rad6 at 2.6 Å resolution. *J. Biol. Chem.*, **273**, 6271–6276.

Yu,H., Peters,J.M., King,R.W., Page,A.M., Hieter,P. and Kirschner,M.W. (1998) Identification of a cullin homology region in a subunit of the anaphase-promoting complex. *Science*, **279**, 1219–1222.

Zachariae,W., Shevchenko,A., Andrews,P.D., Ciosk,R., Galova,M., Stark,J.R., Mann,M. and Nasmyth,K. (1998) Mass spectrometric analysis of the anaphase-promoting complex from yeast: identification of a subunit related to cullins. *Science*, **279**, 1216–1219.

Received July 30, 1999; revised and accepted October 8, 1999

Molecular Biology of the Cell
Vol. 10, 329–344, February 1999

Detection of Transient In Vivo Interactions between Substrate and Transporter during Protein Translocation into the Endoplasmic Reticulum

Martin Dünnwald,* Alexander Varshavsky,† and Nils Johnsson*‡

*Max-Delbrück-Laboratorium, D-50829 Köln, Germany; and †Division of Biology, California Institute of Technology, Pasadena, California 91125

Submitted October 2, 1998; Accepted November 11, 1998
Monitoring Editor: Peter Walter

The split-ubiquitin technique was used to detect transient protein interactions in living cells. N_{ub}, the N-terminal half of ubiquitin (Ub), was fused to Sec62p, a component of the protein translocation machinery in the endoplasmic reticulum of *Saccharomyces cerevisiae*. C_{ub}, the C-terminal half of Ub, was fused to the C terminus of a signal sequence. The reconstitution of a quasi-native Ub structure from the two halves of Ub, and the resulting cleavage by Ub-specific proteases at the C terminus of C_{ub}, serve as a gauge of proximity between the two test proteins linked to N_{ub} and C_{ub}. Using this assay, we show that Sec62p is spatially close to the signal sequence of the prepro-α-factor in vivo. This proximity is confined to the nascent polypeptide chain immediately following the signal sequence. In addition, the extent of proximity depends on the nature of the signal sequence. C_{ub} fusions that bore the signal sequence of invertase resulted in a much lower Ub reconstitution with N_{ub}-Sec62p than otherwise identical test proteins bearing the signal sequence of prepro-α-factor. An inactive derivative of Sec62p failed to interact with signal sequences in this assay. These in vivo findings are consistent with Sec62p being part of a signal sequence-binding complex.

INTRODUCTION

A critical step during the translocation of a protein across the membrane of the endoplasmic reticulum (ER) is the interaction between the signal sequence of a nascent polypeptide and its receptors (Walter *et al.*, 1981; Gilmore and Blobel, 1985; Walter and Johnson, 1994). A stretch of 8 to 12 hydrophobic residues, often at the N terminus of a protein, comprises a signal sequence that is sufficient to initiate the protein's translocation into the endoplasmic reticulum (ER) (Rapoport *et al.*, 1996). To be compatible with a high flux of polypeptides through a limited number of translocation channels in the ER membrane, the interaction between the signal sequence and its receptors has to be short lived. Its transient nature makes such a receptor–ligand interaction difficult to study, especially in living cells. The approaches used for the analysis of protein translocation in cell-free systems circumvent the transience of the signal sequence–receptor interaction by pausing or stopping the synthesis of a nascent polypeptide chain at different stages of its movement to and across the ER membrane (Krieg *et al.*, 1986; Kurzchalia *et al.*, 1986; Connolly *et al.*, 1989). Given these constraints, it is essential to verify in vivo the models derived from in vitro studies. The ability to analyze early translocation events in vivo should also be important for defining the immediate environment of the nascent chain on its path from the ribosome to the ER membrane.

Most of the current methods for detecting protein interactions in vivo either do not operate at the ER membrane or are unable to detect a transient proximity between proteins (Fields and Song, 1989; Aronheim *et al.*, 1997; Rossi *et al.*, 1997; Miyawaki *et al.*, 1997). In the present work, we show that the previously developed split-ubiquitin (split-Ub) technique, also called USPS (Ub/split/protein/sensor) (Johnsson and Varshavsky, 1994a, 1997), is capable of detecting a transient in vivo

---

‡ Corresponding author. E-mail address: johnsson@mpiz-koeln.mpg.de.

© 1999 by The American Society for Cell Biology

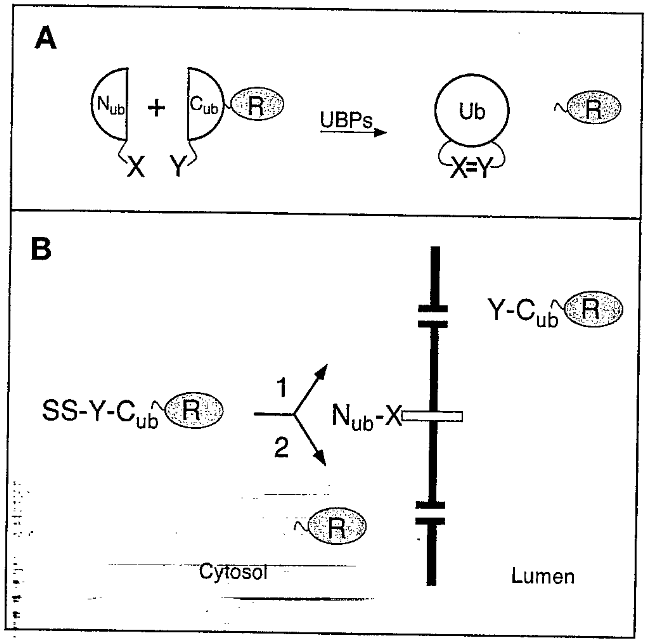


Figure 1. The split-Ub technique and its application to the analysis of protein translocation. (A) N_{ub} and C_{ub} are linked to the interacting proteins X and Y. The X–Y complex brings N_{ub} and C_{ub} into close proximity. N_{ub} and C_{ub} reconstitute a quasi-native Ub moiety, which is cleaved by the UBPs, yielding the free reporter R (Johnsson and Varshavsky, 1994a). (B) Using split-Ub to monitor the proximity between a secretory protein and a component of the translocation machinery. A signal sequence-bearing C_{ub} fusion ($SS\text{-}Y\text{-}C_{ub}\text{-}R$) and a N_{ub} fusion ($N_{ub}\text{-}X$) are coexpressed in a cell. Pathway 1: when N_{ub} is linked to a protein not involved in the targeting for translocation, the uncleaved (except for the signal sequence SS) $Y\text{-}C_{ub}\text{-}R$ enters the ER. Pathway 2: when N_{ub} is linked to a protein involved in the targeting for translocation the signal sequence of the $SS\text{-}Y\text{-}C_{ub}\text{-}R$ brings N_{ub} and C_{ub} into close contact. As a result, some of the $SS\text{-}Y\text{-}C_{ub}\text{-}R$ and $N_{ub}\text{-}X$ molecules interact to form the quasi-native Ub, yielding the free reporter R in the cytosol.

interaction between polypeptides. The split-Ub method is based on the ability of N_{ub} and C_{ub}, the N- and C-terminal halves of Ub, to assemble into a quasi-native Ub. Ub-specific proteases (UBPs), which are present in all eukaryotic cells, recognize the reconstituted Ub, but not its halves, and cleave the Ub moiety off a reporter protein that had been linked to the C terminus of C_{ub}. The liberation of the reporter serves as a readout indicating the reconstitution of Ub. The assay is designed in a way that prevents efficient association of N_{ub} and C_{ub} by themselves, but allows it if the two Ub halves are separately linked to proteins that interact in vivo (Figure 1A). The split-Ub assay has been shown to detect the in vivo dimerization of a leucine zipper-containing domain of the Gcn4p transcriptional activator, and the in vivo interaction between two subunits of the oligosaccharyltransferase complex (Johnsson and Varshavsky, 1994a; Stagljar *et al.*, 1998).

In the present work, we focus on the interaction between Sec62p of the yeast *Saccharomyces cerevisiae* and proteins bearing two different signal sequences. Extensive evidence indicates that Sec62p is a component of the ER translocation machinery (Deshaies and Schekman, 1989; Rothblatt *et al.*, 1989; Müsch *et al.*, 1992). Sec62p is a part of the tetrameric Sec62/63 complex that also contains Sec71p and Sec72p (Deshaies *et al.*, 1991; Feldheim and Schekman, 1994). Sec62/63p can be isolated as a tetramer, or as a part of a larger assembly, the heptameric Sec complex (Panzner *et al.*, 1995). In addition to the Sec62/63 complex, the heptamer contains the trimer of Sec61p. This trimer (Sec61p, Sss1p, Sbh1p in yeast; Sec61α, Sec61β, Sec61γ in mammals) forms the aqueous channel through which a polypeptide chain is translocated across the ER membrane (Simon and Blobel, 1991; Görlich *et al.*, 1992; Crowley *et al.*, 1993, 1994; Mothes *et al.*, 1994; Hanein *et al.*, 1996; Beckmann *et al.*, 1997).

The role of the Sec62/63 tetramer is less well defined. Cross-linking and reconstitution experiments in vitro have shown that Sec62p is close to the nascent polypeptide chain before the initiation of its translocation (Müsch *et al.*, 1992; Lyman and Schekman, 1997; Matlack *et al.*, 1997). One important role of Sec63p is its ability to recruit the Hsp70-type protein Kar2p of the ER lumen to the vicinity of a translocating polypeptide (Brodsky and Schekman, 1993; Lyman and Schekman, 1997). The Sec62/63 complex is essential for the posttranslational translocation of proteins in reconstituted vesicle preparations (Panzner *et al.*, 1995). Genetic analysis supports this conclusion, by showing that the tetrameric Sec62/63 complex is involved in the translocation of proteins whose targeting to the ER membrane is not abolished by the loss of the signal recognition particle (SRP) (Ng *et al.*, 1996). However, it is less clear whether the Sec62/63 complex is the receptor for the signal sequences of those proteins.

In the present work, we demonstrate the ability of the split-Ub assay to detect transient protein interactions in living cells. We show that the assay can monitor a close proximity between Sec62p and a segment of the nascent chain of a signal sequence-bearing protein. The apparent extent of this proximity is influenced by the nature of the signal sequence and the position of C_{ub} in the nascent polypeptide. Our analysis yields a crude map of the environment of the nascent chain during its targeting to and translocation across the ER membrane. Taken together, these findings are the first in vivo evidence that Sec62p, a component of the translocation machinery in the endoplasmic reticulum, is a part of a signal sequence-binding complex.

MATERIALS AND METHODS

Construction of Test Proteins

The C_{ub} fusions 8–13 (Figure 2) were derived from the construct I of Johnsson and Varshavsky (1994a), which encoded Ub-DHFR-ha and

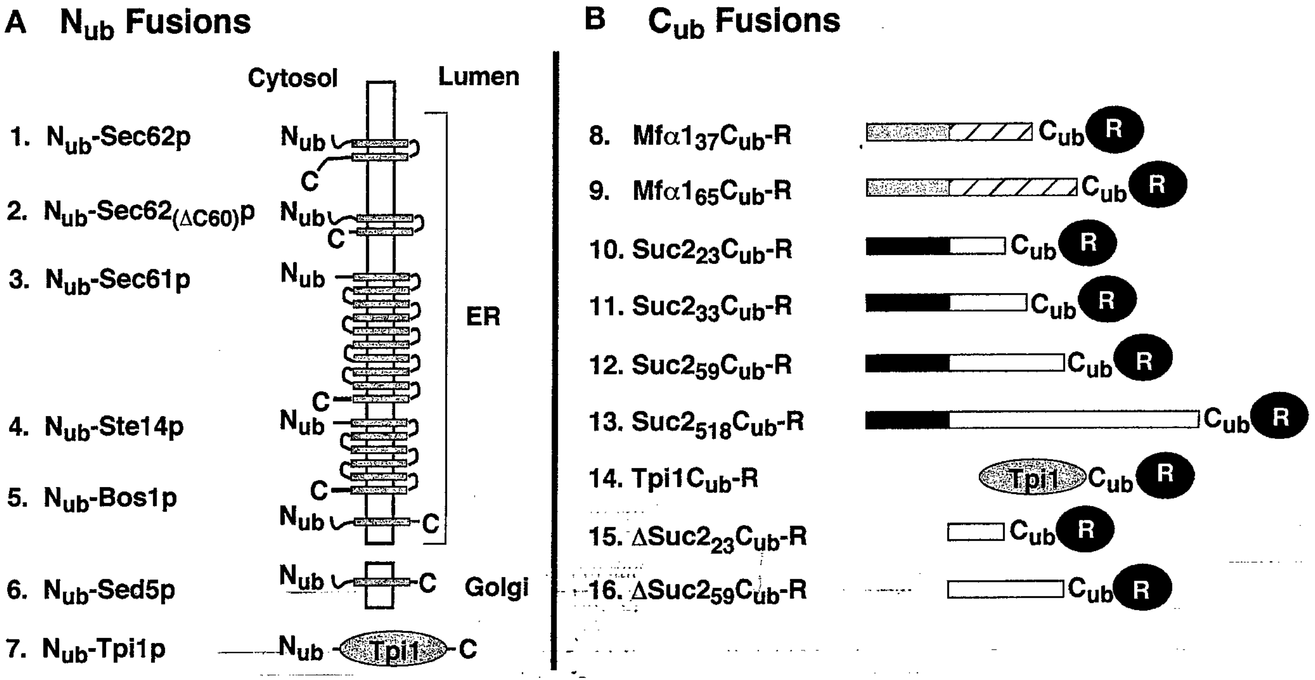


Figure 2. N_{ub} and C_{ub} test fusions. (A) N_{ub} (residues 1–36 of Ub) was fused to the N terminus of either a transmembrane protein (constructs 1–6) or a cytosolic protein (construct 7). The N termini of all proteins are located in the cytosol. The orientations and the numbers of membrane-spanning regions (shaded boxes) were derived from the published studies of these proteins, except for Ste14p, for which the exact number of the domains and the localization of the C terminus are not yet known. The N_{ub} fusions 1–5 are located in the membrane of the ER; the N_{ub} fusion 6 resides in the membrane of the early Golgi. The N_{ub} fusion 2 is a Sec62p derivative lacking the C-terminal 60 residues. The N_{ub} fusion 7 contains the full-length triosephosphate isomerase (N_{ub}-Tpi1p). (B) C_{ub} fusions. The C_{ub} fusions 8 and 9 contain the signal sequence of the prepro-α-factor (shaded boxes), followed by either 37 (construct 8) or 65 residues (construct 9) of the mature α-factor sequence (striped boxes) and a 7-residue linker sequence (not shown). C_{ub} fusions 10–13 contain the signal sequence of the Suc2p invertase (dark boxes) followed by 23 (construct 10), 33 (construct 11), 59 (construct 12), or 518 residues (construct 13) of the mature sequence of invertase (open boxes) and a 7-residue linker sequence (not shown). The C_{ub} fusion 14 contains the complete sequence of *S. cerevisiae* triosephosphate isomerase (Tpi1p) followed by a 17-residue linker peptide and C_{ub}. The C_{ub} fusions 15 and 16 are the signal sequence-lacking counterparts of the fusions 10 and 12. C_{ub} is always followed by a reporter protein. The reporter is DHFR-ha or Ura3p for the C_{ub} fusions 8–13, and DHFR-ha for the C_{ub} fusions 14, 15, and 16.

contained a *Bam*HI site at the amino acid position 36 of Ub, and from the previously described Ub translocation constructs I, VI, IX, X, XXIII, and XXV (Johnsson and Varshavsky, 1994b). The above *Bam*HI site of the Ub-DHFR-ha construct I was fused to a linker sequence in which a 5′-*Sal*I site allowed the in-frame insertion of an *Eag*I–*Sal*I fragment containing the promoter, the signal sequence, and a portion of the mature sequence of the corresponding Ub fusions. The newly introduced sequence was G TCG ACC ATG TCG GGG GGG ATC CCT. The last three triplets encode residues 35, 36, and 37 of Ub (the beginning of C_{ub}). The underlined sequences are the *Sal*I and *Bam*HI sites, respectively. The final constructs were in the single-copy plasmids pRS314 or pRS315 (Sikorski and Hieter, 1989). Expression of the C_{ub} fusions bearing Dha (DHFR-ha) as a reporter was mediated by the P_{ADH1} promoter, except for the C_{ub} fusion 14, which was expressed from the P_{CUP1} promoter. The same promoter was used for expressing the Ura3p-based C_{ub} fusions.

The C_{ub} fusions 15 and 16 (Figure 2) were derived from constructs 10 and 12 by deleting the *Hin*dIII fragment spanning the first four codons of the *SUC2* ORF and a short portion of the polylinker sequence between the 3′-end of the P_{ADH1} promoter and the *SUC2* ORF. As a result, the translation of C_{ub} fusions 15 and 16 began at the first codon of the mature invertase, skipping its signal sequence. The C_{ub} fusion 14 (Figure 2) was produced through an in-frame fusion of a PCR fragment containing the complete *TPI1* coding sequence and C_{ub}-Dha. The sequence between *TPI1* and C_{ub} is as follows: AAC GGG TCG ACC GAC TAC AAG GAC GAC GAT GAC AAG GGC TCG ACC ATGTCG GGG GGG ATC CCT. The underlined sequences indicate, respectively, the last codon of *TPI1* and the first three codons of C_{ub}.

A fragment encoding N_{ub}-Sec62p was constructed using PCR amplification of a 1050 base pair (bp) fragment containing the *SEC62* ORF. PCR introduced a *Bam*HI site and a linker sequence in front of the start codon of *SEC62* and an *Xho*I site 173 bp downstream of the stop codon. The P_{CUP1}-N_{ub} modules were cloned as *Bam*HI fragments in frame with the *SEC62* ORF. The sequence between N_{ub} and *SEC62* is GGG ATC CCT TCT GGG ATG. The first three codons encode residues 35, 36, and 37 of N_{ub}, followed by the Gly-Ser linker and the start codon of *SEC62*. The *Bam*HI site is underlined. The final constructs resided in pRS316 or pRS313. N_{ub}-*TPI1*, N_{ub}-*SED5*, N_{ub}-*STE14* (a gift from N. Lewke), and N_{ub}-Sec62(ΔC60)-Dha were constructed similarly to N_{ub}-*SEC62*. With the exception of N_{ub}-Sec62(ΔC60)-Dha, which was placed in pRS316 and pRS313, all of these fusions resided in pRS314. The linker connecting codon 35 of N_{ub} and the first codon of a linked gene was GGG ATC CCT GGG GAT ATG for N_{ub}-*TPI1* and N_{ub}-*SED5*, and GGG ATC CCT GGG GAT CAC for N_{ub}-*STE14*. Underlined are the *Bam*HI site and the first codon of the linked gene. The sequence GGG TCG ACC TTA ATG CAG AGA TCT GGC ATC ATG GTT connected the last codon

Table 1. Yeast strains

Strain	Relevant genotype	Source/comment
YPH500	*MAT*α *ade2-101 his3-*Δ*200 leu2-*Δ*1 lys2-801 trp1-*Δ*63 ura3-52*	Sikorski and Hieter (1989)
JD53	*MAT*α *his3-*Δ*200 leu2-3,112 lys2-801 trp1-*Δ*63 ura3-52*	Dohmen *et al.* (1995)
RSY529	*MAT*α *his4 leu2-3,112 ura3-52 sec62-1*	Deshaies and Schekman (1989)
NJY51	*MAT*α *ade2-101 his3-*Δ*200 leu2-*Δ*1 lys2-801 trp1-*Δ*63 ura3-52* *NUB-BOS1::pRS306*	Derivative of YPH500
NJY62-1	*MAT*α *his3-*Δ*200 leu2-3,112 lys2-801 trp1-*Δ*63 ura3-52* *NUB-SEC62::pRS306*	Derivative of JD53
NJY73-I	*MAT*α *his3-*Δ*200 leu2-3,112 lys2-801 trp1-*Δ*63 ura3-52* *NUB-BOS1::pRS303*	Derivative of JD53
NJY74-I	*MAT*α *his3-*Δ*200 leu2-3,112 lys2-801 trp1-*Δ*63 ura3-52* *NUB-BOS1::pRS306*	Derivative of JD53
NJY61-I	*MAT*α *his3-*Δ*200 leu2-3,112 lys2-801 trp1-*Δ*63 ura3-52* *NUB-SEC61::pRS304*	Derivative of JD53
NJY61-A	*MAT*α *his3-*Δ*200 leu2-3,112 lys2-801 trp1-*Δ*63 ura3-52* *NUA-SEC61::pRS304*	Derivative of JD53
NJY61-G	*MAT*α *his3-*Δ*200 leu2-3,112 lys2-801 trp1-*Δ*63 ura3-52* *NUG-SEC61::pRS304*	Derivative of JD53

of *SEC62* in N_{ub}-Sec62(ΔC60)-Dha (codon 223, underlined) to the first two codons of DHFR (underlined).

N_{ub}-*BOS1* was constructed in part by PCR amplification, with two synthetic oligos and yeast genomic DNA as a template, yielding a 258-bp fragment containing the first 229 bp of the *BOS1* ORF. Upstream of the *BOS1* ATG was a short linker sequence and a *Bam*HI site, to allow in-frame fusion of the P_{CUP1} promoter-N_{ub} module. The sequence between N_{ub} and *BOS1* reads: GGG ATC CCT CCA GGA ATG. The first four triplets encode residues 35, 36, 37, and 38 of N_{ub}, followed by the Gly codon and the start codon of *BOS1*. The *Bam*HI site is underlined. The 3′-region of the resulting fragment terminated in a *Sal*I site for insertion into the integrating vectors pRS306 or pRS303. The vector was cut at the unique *Eco*RI site in the *BOS1*-containing fragment and transformed into *S. cerevisiae* strains YPH500 and JD53 to produce, through homologous recombination, the integrated cassette that expressed N_{ub}-Bos1p from the P_{CUP1} promoter. The presence of the desired gene fusion and the absence of wild-type *BOS1* were verified by PCR.

An integrated copy of P_{CUP1}-N_{ub}-*SEC62* was produced by amplifying the first 438 bp of the *SEC62* ORF, and then cloning it, using the *Bam*HI and *Eco*RI restriction sites, in frame behind the pRS306-P_{CUP1}-N_{ub} cassette. A unique *Afl*II site in the *SEC62* ORF was used to linearize the plasmid for transformation and integration at the *S. cerevisiae SEC62* gene, yielding the strain NJY62-I. The N-terminal 147-residue fragment of Sec62p that was coexpressed with N_{ub}-Sec62p in the resulting strain has previously been shown to be inactive in translocation (Deshaies and Schekman, 1990). N_{ub}-*SEC61* was constructed by targeted integration of a N_{ub}-*SEC61*–containing fragment into *SEC61* of the *S. cerevisiae* strain JD53 (Table 1). Specifically, a fragment containing the first 875 bp of the *SEC61* ORF was amplified by PCR and inserted downstream of the pRS304- or pRS303-based P_{CUP1}-N_{ub} cassette, using the flanking *Bam*HI and *Eco*RI sites. The linker sequence between N_{ub} and *SEC61* was GGG ATC CCT GGG TCT GGG ATG. Underlined are the *Bam*HI site and the first codon of *SEC61*. For targeted integration, the plasmid was linearized at the unique *Stu*I site in the *SEC61* ORF to create the yeast NJY61-I. A detailed description of the NJY61 strains (Table 1) will be presented elsewhere (Wittke and Johnsson, unpublished data).

All of the P_{CUP1} promoter-controlled ORFs were expressed under noninducing conditions (no copper added to the medium), except in the experiment shown in Figure 5B, where cells were incubated in the presence of 0.1 mM $CuSO_4$.

Immunoblotting

Proteins fractionated by SDS-12.5% PAGE were electroblotted onto nitrocellulose (Schleicher & Schuell, Dassel, Germany) or polyvinylidene difluoride (Machery-Nagel, Düren, Germany) membranes, using the semidry transfer system (Hoeffer Pharmacia Biotech, San Francisco, CA). Blots were incubated with an anti-ha monoclonal antibody (Babco, Richmond, CA) and visualized using horseradish peroxidase-coupled goat anti-mouse antibody (Bio-Rad, Hercules, CA), the chemiluminescence detection system (Boehringer, Mannheim, Germany), and x-ray films. Where indicated, quantification was performed using the Lumi Imager system (Boehringer).

Pulse-Chase Analysis

Yeast-rich (YPD) and synthetic minimal media with 2% dextrose (SD) were prepared as described previously (Dohmen *et al.*, 1995). *S. cerevisiae* cells expressing the N_{ub} and C_{ub} fusions were grown at 30°C in 10 ml of SD medium without externally added copper to an OD_{600} of ~1 and labeled for 5 min with Redivue Promix-[^{35}S] (Amersham, Buckinghamshire, United Kingdom), followed (either directly or after a chase) by immunoprecipitation with the anti-ha monoclonal antibody, essentially as described by Johnsson and Varshavsky (1994a,b). The EndoH analysis of glycosylated proteins was carried out as described by Orlean *et al.* (1991). Samples were concentrated before SDS-12.5% PAGE by precipitation with chloroform/methanol. Gels were fixed and enhanced for fluorography. For quantitative analysis, a dried gel was exposed and scanned using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA).

RESULTS

Experimental Strategy

The use of split Ub to monitor the proximity between the proteins X and Y requires the construction of two "complementary" fusion proteins. One fusion bears N_{ub} (see INTRODUCTION) linked to X (N_{ub}-X) and the other bears C_{ub} linked to both Y and a reporter protein R at the C terminus of C_{ub} (Y-C_{ub}-R). The liberation of the reporter through the Ub-dependent cleavage by UBPs indicates the in vivo reconstitution of a quasi-native Ub from N_{ub} and C_{ub}. In the split-Ub assay, the efficiency of cleavage at the C terminus of C_{ub} in Y-C_{ub}-R is measured relative to the efficiency of cleavage observed with selected reference (control) proteins (Figure 1).

To monitor protein interactions during translocation of a protein across the ER membrane, N_{ub} was fused to the N terminus of a membrane protein that is a part of the translocation machinery (Figure 1). Owing to the constraint of the assay, which requires the cytosolic location of the reconstituted Ub, the N terminus of this membrane protein must be located in the cell's cytosol. Sec62p has an N-terminal cytosolic domain of 158 residues, which is followed by two membrane-spanning segments and a C-terminal segment also facing the cytosol (Deshaies and Schekman, 1990). N_{ub} was therefore fused to the N terminus of Sec62p, yielding N_{ub}-Sec62p. C_{ub} was sandwiched between the 56 N-terminal residues of the precursor of *S. cerevisiae* α-factor pheromone (prepro-α-factor) and the ha epitope-tagged mouse dihydrofolate reductase (DHFR-ha; denoted as Dha) as a reporter protein, yielding Mfα_{37}-C_{ub}-Dha (Figure 2). The cleavage of the C_{ub}-containing fusion at the C_{ub}–Dha junction was detected with a monoclonal anti-ha antibody.

Split-Ub Detects a Proximity between a Translocating Protein and Sec62p

We first verified that Mfα_{37}-C_{ub}-Dha could be translocated across the ER membrane and that the N-terminal extension of Sec62p with N_{ub} did not interfere with the Sec62p function in translocation. After a 5-min pulse of wild-type *S. cerevisiae* with ^{35}S-methionine, the labeled Mfα_{37}-C_{ub}-Dha was immunoprecipitated as a glycosylated and unclipped fusion (Figure 3A). Thus, Mfα_{37}-C_{ub}-Dha could indeed be translocated into the lumen of ER. Introduction of the same Mfα_{37}-C_{ub}-Dha construct into the yeast strain RSY529, which carries a temperature-sensitive (ts) variant of Sec62p (Rothblatt *et al.*, 1989), confirmed the severe translocation defect of this strain. About 50% of the pulse-labeled Mfα_{37}-C_{ub}-Dha entered the lumen of the ER in this strain at the semipermissive temperature of 30°C, while the rest remained in the cytosol (Figure 3A). Thus, the translocation of Mfα_{37}-C_{ub}-Dha depends on Sec62p. This made it possible to determine whether N_{ub}-Sec62p is functionally active. The test utilized N_{ug}-Sec62p, in which the N-terminal half of Ub contained Gly-13 instead of wild-type Ile-13. This derivative, denoted as N_{ug}, has a lower affinity for C_{ub} than the wild-type N_{ub} (Johnsson and Varshavsky, 1994a). We chose N_{ug}-Sec62p for this experiment to minimize the reconstitution of the Ub moiety through interactions between N_{ub}-Sec62p and potentially arrested molecules of Mfα_{37}-C_{ub}-Dha, which might be localized in the cytosol. Plasmids expressing N_{ug}-Sec62p and Mfα_{37}-C_{ub}-Dha were cotransformed into RSY529 cells and assayed at 30°C. As in wild-type cells, only translocated Mfα_{37}-C_{ub}-Dha, but virtually no free Dha or nontranslocated Mfα_{37}-C_{ub}-Dha, was detected after immunoprecipitation and EndoH treatment of the

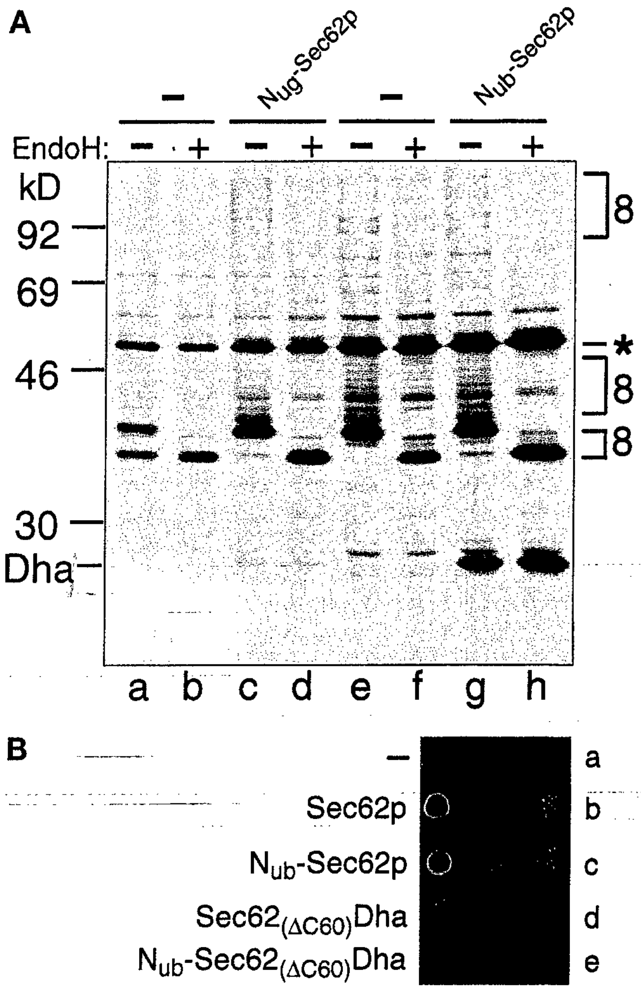


Figure 3. Sec62p is close to the signal sequence of the α-factor precursor. (A) *S. cerevisiae* cells expressing Mfα_{37}-C_{ub}-Dha (construct 8; Figure 2) were labeled for 5 min with ^{35}S-methionine. The extracted proteins were immunoprecipitated with anti-ha antibody, followed by a mock treatment (lanes a, c, e, and g) or the treatment with EndoH (lanes b, d, f, and h), and SDS-PAGE. The results with cells coexpressing N_{ug}- or N_{ub}-Sec62p are shown in lanes c and d and g and h, respectively. The analysis was performed with N_{ug}-Sec62p in the *S. cerevisiae* mutant RSY529 carrying a ts allele of *SEC62* (lanes a–d) or with N_{ub}-Sec62p in the wild-type yeast (lanes e–h). Number 8 (following the numbering of the constructs in Figure 2) on the right indicates the positions of uncleaved Mfα_{37}-C_{ub}-Dha and its glycosylated forms. An asterisk denotes an unrelated yeast protein that cross-reacts with the anti-ha antibody. (B) N_{ub}-Sec62p encodes a functionally active protein. RSY529 cells carrying an empty plasmid (a), Sec62p (b), N_{ub}-Sec62p (c), Sec62(ΔC60)Dha (d), or N_{ub}-Sec62(ΔC60)Dha (e) were spotted on minimal media and grown for 2 d at 30°C (semipermissive temperature for unmodified RSY529).

cells that had been labeled for 5 min with ^{35}S-methionine (Figure 3A).

To test N$_{ub}$-Sec62p directly (N$_{ub}$ is the wild-type half of Ub, containing Ile at position 13), we examined its ability to complement the growth defect of RSY529 cells. RSY529 cells expressing N$_{ub}$-Sec62p were found to grow at the semipermissive temperature of 30°C, in contrast to congenic cells carrying a control plasmid (Figure 3B). To verify that the suppression of the ts phenotype was not due to the initiation of translation from the first (internal) ATG codon of Sec62p within the N$_{ub}$-Sec62p fusion, the rescue experiment was successfully repeated with the otherwise identical derivative of N$_{ub}$-Sec62p that lacked the first ATG of *SEC62* (our unpublished results).

A significant amount of free Dha was generated when Mfα_{37}-C$_{ub}$-Dha was expressed (in either wild-type or RSY529 cells) together with N$_{ub}$-Sec62p, which contained the wild-type half of Ub (Figure 3A and our unpublished results). We concluded that Sec62p is close to the nascent polypeptide chain during its translocation into the ER. The cleavage at the C terminus of C$_{ub}$ requires its interaction with N$_{ub}$ and depends on the presence of UBPs (Johnsson and Varshavsky, 1994a). Since UBPs have previously been shown to be absent from the ER (Johnsson and Varshavsky, 1994b), the free Dha moiety had to be produced in the cytosol. Fractionation experiments confirmed that free Dha was absent from membrane-enclosed compartments in whole-cell extracts (our unpublished results). An entirely independent evidence for this conclusion was produced by replacing Dha in Mfα_{37}-C$_{ub}$-Dha with Ura3p as the reporter moiety. Ura3p confers the Ura$^+$ phenotype on *ura3Δ* cells only if Ura3p has access to the cytosol (Johnsson and Varshavsky, 1994b). In our tests, the cytosolic Ura3p was produced only if Mfα_{37}-C$_{ub}$-Ura3p was coexpressed with N$_{ub}$-Sec62p (compare A and B in Figure 7), in agreement with the other evidence (see above) that the cleavage at the C$_{ub}$–protein junction takes place exclusively in the cytosol.

The transient nature of the proximity between Sec62p and the nascent chain of a translocated protein was indicated by the near-absence of the released Dha moiety if Mfα_{37}-C$_{ub}$-Dha was coexpressed with either N$_{ug}$-Sec62p or N$_{ua}$-Sec62p instead of N$_{ub}$-Sec62p (N$_{ua}$ denotes Ala at position 13 of N$_{ub}$); by contrast, the same experiment with N$_{ub}$-Sec62p resulted in a significant cleavage of Mfα_{37}-C$_{ub}$-Dha (Figures 3A and 6C). Previous work (Johnsson and Varshavsky, 1994a) has shown that N$_{ua}$ and N$_{ug}$ can induce significant Ub reconstitution when either of them and C$_{ub}$ are linked to polypeptides that form a stable (long-lived) complex in a cell. In summary, the observed absence of significant Ub reconstitution with N$_{ua}$ and N$_{ug}$ (in contrast to N$_{ub}$) was interpreted to signify a close but transient (short-lived) proximity between Sec62p and Mfα_{37}-C$_{ub}$-Dha.

Specificity of the Spatial Proximity between a Signal Sequence-bearing Nascent Polypeptide and Sec62p

A commonly used negative control in a translocation assay is a protein with a defective or absent signal sequence (Allison and Young, 1988; Müsch *et al.*, 1992). Such a control is not entirely compatible with spatio-temporal aspects of the split-Ub assay. Specifically, a C$_{ub}$-fusion protein lacking a signal sequence accumulates in the cytosol (where the split-Ub assay operates), whereas an analogous signal sequence-bearing protein is continuously removed from this compartment. A direct comparison between reactions that involve a signal sequence-bearing polypeptide and its signal sequence-lacking counterpart requires the ability to compare the local concentrations of the two polypeptides at the site of translocation. We are not aware of an in vivo technique that would be independent of the split-Ub assay and at the same time would allow a measurement of these parameters. Therefore, we devised an alternative control. The extent of cleavage of Mfα_{37}-C$_{ub}$-Dha at the C$_{ub}$–Dha junction should reflect the time-averaged spatial proximity between the nascent Mfα-chain and a coexpressed N$_{ub}$-containing fusion. By comparing the extent of cleavage of Mfα_{37}-C$_{ub}$-Dha in the presence of N$_{ub}$-Sec62p (Figure 3A) with the analogous activity of N$_{ub}$-fusion proteins that are not involved in the ER targeting and translocation, we could assess the specificity of the reaction between N$_{ub}$-Sec62p and Mfα_{37}-C$_{ub}$-Dha.

Four N$_{ub}$-fusion proteins, N$_{ub}$-Bos1p, N$_{ub}$-Ste14p, N$_{ub}$-Sed5p, and N$_{ub}$-Tpi1p were tested in the split-Ub assay with Mfα_{37}-C$_{ub}$-Dha. The expected intracellular locations of these N$_{ub}$ fusions, and their predicted topologies in the membrane are shown in Figure 2A (Shim *et al.*, 1991; Banfield *et al.*, 1994; Sapperstein *et al.*, 1994; Lewke and Johnsson, unpublished data). We found that, in contrast to N$_{ub}$-Sec62p, none of the four tested N$_{ub}$ fusions induced a significant cleavage of Mfα_{37}-C$_{ub}$-Dha (Figure 4A). A small amount of free Dha could be detected in the immunoblots when N$_{ub}$-Bos1p was overexpressed. The lack of significant Ub reconstitution from N$_{ub}$ and C$_{ub}$ upon coexpression of Mfα_{37}-C$_{ub}$-Dha and the N$_{ub}$-modified ER membrane proteins, Bos1p, Ste14p (Figure 4A), and Sec12p (Nakano *et al.*, 1988; our unpublished results), confirmed that the steady-state concentration of Mfα_{37}-C$_{ub}$-Dha in the cytosol was extremely low.

To verify that the observed absence of Ub reconstitution (Figure 4A) was not due to either low concentrations of the tested fusion proteins or reduced accessibility of their linked N$_{ub}$ moieties, we compared the activity of these N$_{ub}$ fusions toward a cytosolic C$_{ub}$-fusion protein. C$_{ub}$-Dha was fused to the C terminus of the cytosolic enzyme triosephosphate isomerase

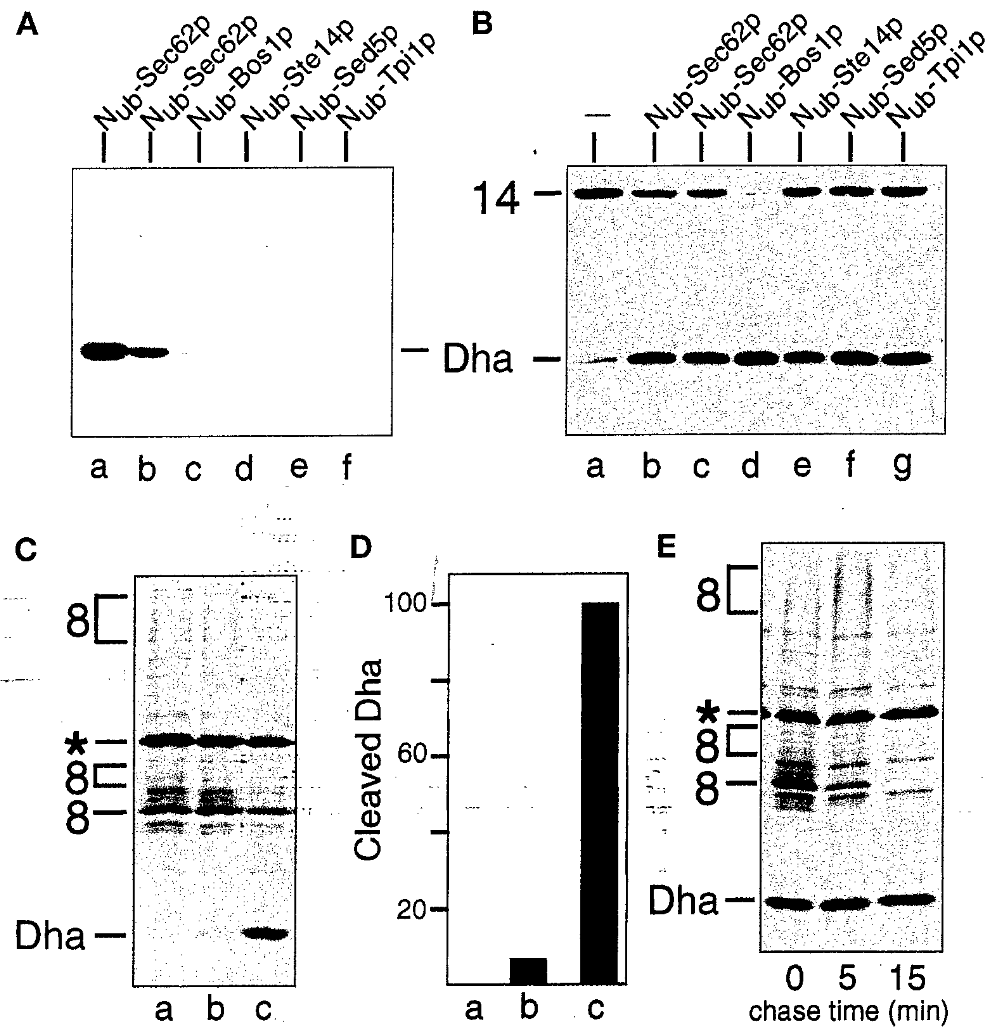


Figure 4. The in vivo proximity between Sec62p and Mfα_{37}-C_{ub}-Dha is transient and specific. (A) Immunoblot analysis of extracts of *S. cerevisiae* coexpressing the Mfα_{37}-C_{ub}-Dha (construct 8; Figure 2) and one of the following constructs: N_{ub}-Sec62p, integrated (lane a) or plasmid-borne (lane b); N_{ub}-Bos1p (lane c); N_{ub}-Ste14p (lane d); N_{ub}-Sed5p(lane e); and N_{ub}-Tpi1p (lane f). (B) Immunoblot analysis of extracts of *S. cerevisiae* expressing the Tpi1p-C_{ub}-Dha fusion (construct 14; Figure 2) alone (lane a) or together with one of the following constructs: N_{ub}-Sec62p, either integrated (lane b) or plasmid-borne (lane c); N_{ub}-Bos1p (lane d); N_{ub}-Ste14p (lane e); N_{ub}-Sed5p(lane f); and N_{ub}-Tpi1p (lane g). Number 14 on the left indicates the position of uncleaved Tpi1p-C_{ub}-Dha. (C) *S. cerevisiae* cells expressing Mfα_{37}-C_{ub}-Dha (construct 8; Figure 2) together with either the vector (lane a), N_{ub}-Bos1p (lane b), or N_{ub}-Sec62p (lane c) were labeled for 5 min with ^{35}S-methionine. The extracted proteins were immunoprecipitated with anti-ha antibody and analyzed by SDS-PAGE. (D) Quantitation of the pulse-labeling experiment (C) using PhosphorImager. The extent of Dha release in the presence of N_{ub}-Sec62p was arbitrarily set at 100. The averages of three experiments are shown. Lanes a, b, and c are the same as in panel C. (E) *S. cerevisiae* cells expressing Mfα_{37}-C_{ub}-Dha together with N_{ub}-Sec62p were labeled for 5 min with ^{35}S-methionine and chased for 5 and 15 min, followed by extraction of proteins, immunoprecipitation with anti-ha antibody, and SDS-PAGE.

(Tpi1p), yielding Tpi1-C_{ub}-Dha (Figure 2B). All of the N_{ub}-fusion proteins in Figure 2A induced a significant release of Dha from the test protein Tpi1-C_{ub}-Dha (Figure 4B). This analysis also suggested that N_{ub}-Bos1p was expressed to higher levels than other N_{ub} fusions.

To quantify the relative proximities of N_{ub}-Bos1p and N_{ub}-Sec62p to Mfα_{37}-C_{ub}-Dha, yeast cells were labeled for 5 min with ^{35}S-methionine, and the released Dha was determined as described in the legend to Figure 4. Coexpression of N_{ub}-Sec62p and Mfα_{37}-C_{ub}-Dha yielded ~15 times more of the free Dha than coexpression of N_{ub}-Bos1p and Mfα_{37}-C_{ub}-Dha (Figure 4, C and D). Assuming that the N_{ub} moieties in N_{ub}-Sec62p and N_{ub}-Bos1p were equally accessible to the cytosol (Figure 4B), we concluded that the time-averaged proximity between the nascent chain of Mfα_{37}-C_{ub}-Dha and the N_{ub}-bearing transmembrane proteins was much higher for Sec62p than for the ER membrane proteins that are not involved in targeting or translocation. Note that this analysis may actually underestimate the proximity of Sec62p to the nascent chain, because we invariably observed a more efficient cleavage of Mfα_{37}-C_{ub}-Dha when N_{ub}-Sec62p was the only form of Sec62p in the cell (Figures 4A and 6A). Therefore we interpret the reduced cleavage of Mfα_{37}-C_{ub}-Dha in the presence of both N_{ub}-Sec62p and the native Sec62p as the consequence of competition between those two Sec62p-containing species for either the signal sequences of translocated proteins or the ligands of Sec62p in the complex of Sec proteins.

Recent evidence indicates that misfolded or otherwise abnormal proteins in the lumen of the ER can be retrotransported across the ER membrane back into

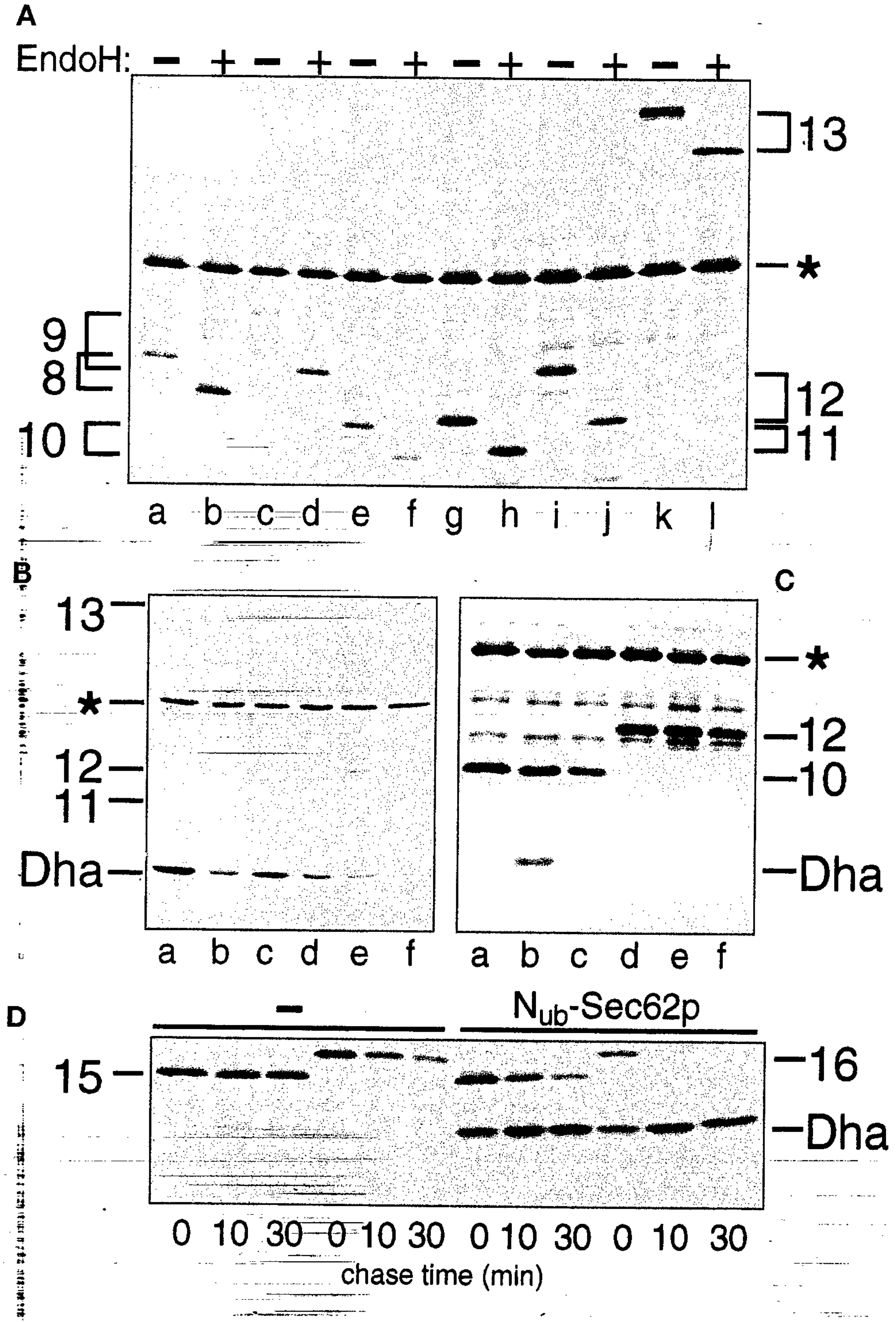


Figure 5. The nature of the signal sequence and its distance from C_{ub} determine the extent of cleavage of C_{ub}-R in the presence of N_{ub}-Sec62p. (A) *S. cerevisiae* expressing $Mf\alpha_{37}$-C_{ub}-Dha (construct 8; Figure 2) (lanes a and b), $Mf\alpha_{65}$-C_{ub}-Dha (construct 9) (lanes c and d), $Suc2_{23}$-C_{ub}-Dha (construct 10) (lanes e and f), $Suc2_{33}$-C_{ub}-Dha (construct 11) (lanes g and h), $Suc2_{59}$-C_{ub}-Dha (construct 12) (lanes i and j) and $Suc2_{518}$-C_{ub}-Dha (construct 13) (lanes k and l) were labeled with ^{35}S-methionine for 5 min. The extracted proteins were either mock-treated (lanes a, c, e, g, i, and k) or treated with EndoH (lanes b, d, f, h, j, and l), followed by immunoprecipitation with anti-ha antibody and SDS-PAGE. (B) Same as panel A but the cells also contained N_{ub}-Sec62p in addition to the Cub-fusions $Mf\alpha_{37}$-C_{ub}-Dha, $Mf\alpha_{65}$-C_{ub}-Dha, $Suc2_{23}$-C_{ub}-Dha, $Suc2_{33}$-C_{ub}-Dha, $Suc2_{59}$-C_{ub}-Dha, $Suc2_{518}$-C_{ub}-Dha (lanes a–f). The analysis was carried out by immunoblotting whole- cell extracts with the anti-ha antibody. (C) *S. cerevisiae* cells expressing $Suc2_{23}$-C_{ub}-Dha (construct 10; Figure 2) (lanes a–c) and $Suc2_{59}$-C_{ub}-Dha (construct 12; Figure 2) (lanes d–f) together with either N_{ub}-Sec62p (lanes b and e), N_{ub}-Bos1p (lanes c and f) or the vector (lanes a and d) were labeled for 5 min with ^{35}S-methionine. Whole-cell extracts were immunoprecipitated with anti-ha antibody, followed by SDS-PAGE and autoradiography. (D) *S. cerevisiae* cells expressing $\Delta Suc2_{23}$-C_{ub}-Dha (construct 15; Figure 2) or $\Delta Suc2_{59}$-C_{ub}-Dha (construct 16; Figure 2) together with either the vector (first six lanes) or N_{ub}-Sec62p (last six lanes) were labeled for 5 min with ^{35}S-methionine and chased for 10 and 30 min, followed by extraction, immunoprecipitation with anti-ha antibody, and SDS-PAGE. Numbers 15 and 16 indicate the positions of the corresponding (uncleaved) C_{ub} fusions.

the cytosol, where they are degraded by the Ub system (Biederer *et al.*, 1996; Hiller *et al.*, 1996; Wiertz *et al.*, 1996). This retrotransport involves at least some of the components of the known ER translocation machinery (Plemper *et al.*, 1997). To determine whether the cleavage of $Mf\alpha_{37}$-C_{ub}-Dha at the C_{ub}-Dha junction occurs during translocation into the ER or during (in this case) a hypothetical retrotransport from the ER, cells coexpressing N_{ub}-Sec62p and $Mf\alpha_{37}$-C_{ub}-Dha were labeled for 5 min with ^{35}S-methionine, and then chased for 15 min (Figure 4E). Although the translocated $Mf\alpha_{37}$-C_{ub}-Dha disappeared rapidly during the chase, the amount of free Dha that accumulated during the pulse remained constant.

We conclude that the in vivo proximity between Sec62p and Mfα_{37}-C_{ub}-Dha that is detected by the split-Ub assay occurs either during or very shortly after the synthesis of Mfα_{37}-C_{ub}-Dha. The apparent disappearance of the pulse-labeled, translocated Mfα_{37}-C_{ub}-Dha during the chase accounts for the difficulty in detecting this species by a steady-state assay such as immunoblotting (Figures 4A, 5B, and 6A). The likely cause of the disappearance of translocated Mfα_{37}-C_{ub}-Dha is its molecular mass heterogeneity, owing to its glycosylation, which results in a smear upon SDS-PAGE (Figures 3A, 4C, and 4E).

The Efficiency of Ub Reconstitution Mediated by N_{ub}-Sec62p Depends on Both the Identity of a Signal Sequence and the Position of C_{ub} in the Nascent Polypeptide Chain

The proximity of Sec62p to the signal sequence of Mfα_{37}-C_{ub}-Dha is detected, in the split-Ub assay, through the ability of N_{ub}-Sec62p to induce the cleavage of Mfα_{37}-C_{ub}-Dha at the C_{ub}–Dha junction (Figure 3A). If this cleavage reflects the physical proximity between Sec62p and a signal sequence, the efficiency of cleavage should decrease if the C_{ub} moiety is moved closer to the C terminus of the nascent polypeptide chain. However, this purely spatial consideration neglects the temporal aspect of the translocation process (Walter and Johnson, 1994). The targeting and the actual translocation are initiated during or shortly after the synthesis of a signal sequence-bearing protein. Consequently, the C-terminal parts of the nascent chain may still be synthesized, or at least associated with the ribosome, at the time when Sec62p and the signal sequence have already become spatially close. Extending the spacer would increase the distance between C_{ub} and the signal sequence of Mfα_{37}-C_{ub}-Dha. This would be expected to decrease the time window available for the interaction between the C_{ub} moiety of Mfα_{37}-C_{ub}-Dha and the N_{ub} moiety of N_{ub}-Sec62p. Therefore, a test of this kind cannot deconvolute the contribution of each of the two parameters (increased spatial distance along the chain between Sec62p and C_{ub} and decreased time window for the N_{ub}–C_{ub} interaction) to the overall effect of extending the length of the polypeptide between the signal sequence and C_{ub}. These constraints notwithstanding, moving the C_{ub} moiety of Mfα_{37}-C_{ub}-Dha further away from its signal sequence makes it possible to gauge the accessibility of Sec62p to specific regions of the nascent polypeptide chain in vivo.

In the actual experiment, the distance between the signal sequence of Mfα_{37}-C_{ub}-Dha and its C_{ub} moiety was increased from 37 to 65 residues (Mfα_{65}-C_{ub}-Dha; Figure 2B, construct 9). The results of EndoH treatment of Mfα_{65}-C_{ub}-Dha immunoprecipitated from pulse-labeled wild-type cells confirmed that Mfα_{65}-C_{ub}-Dha was efficiently translocated into the ER (Figure 5A). However, the efficiency of the Dha-yielding cleavage of Mfα_{65}-C_{ub}-Dha upon coexpression of N_{ub}-Sec62p was clearly reduced in comparison to the same cleavage with Mfα_{37}-C_{ub}-Dha and N_{ub}-Sec62p (Figure 5B).

Both the kinetics and the mode of targeting for translocation are influenced by the identity of a signal sequence (Bird *et al.*, 1987; Johnsson and Varshavsky, 1994b; Ng *et al.*, 1996). For example, the efficient translocation of invertase (Suc2p) requires the SRP, in contrast to a much weaker requirement for SRP in the case of the prepro-α-factor's signal sequence (Hann and Walter, 1991; Ogg *et al.*, 1992; Johnsson and Varshavsky, 1994b). Consequently, the coupling between translation and translocation is tighter for proteins bearing the invertase signal sequence than for proteins carrying the signal sequence of the α-factor.

We assessed the in vivo proximity of the invertase signal sequence to Sec62p by measuring the reconstitution of Ub from N_{ub}-Sec62p and Suc2-C_{ub}-Dha, where the Suc2p moiety was linked to C_{ub} through a spacer of increasing length (Figure 2B). The expression and efficient translocation of different Suc2-C_{ub}-Dha constructs were assayed by immunoprecipitation and subsequent EndoH treatment (Figure 5A). The proximity of C_{ub} in Suc2-C_{ub}-Dha to N_{ub} of N_{ub}-Sec62p was assayed by immunoblot detection of the cleavage-derived free Dha in whole-cell extracts. The pattern already observed for the Mfα-C_{ub}-Dha constructs recurred with the constructs bearing the invertase signal sequence (Figure 5B). Moreover, coexpression of N_{ub}-Sec62p with either Suc2$_{23}$-C_{ub}-Dha or Suc2$_{33}$-C_{ub}-Dha yielded lower amounts of free Dha than the analogous assays with N_{ub}-Sec62p and Mfα_{37}-C_{ub}-Dha, which bears a spacer of comparable length (Figure 5B).

Pulse-chase analyses with cells expressing N_{ub}-Sec62p (or N_{ub}-Bos1p) and either Suc2$_{23}$-C_{ub}-Dha or Suc2$_{59}$-C_{ub}-Dha confirmed the immunoblot data. Specifically, a significant release of free Dha was observed only for the pair of N_{ub}-Sec62p and Suc2$_{23}$-C_{ub}-Dha (Figure 5C). Our previous work has shown that the segment of the nascent polypeptide chain where the C_{ub} moiety was inserted in either the Suc2$_{23}$-C_{ub}-Dha or the Mfα_{37}-C_{ub}-Dha fusion is transiently exposed to the cytosol—until the initiation of ER translocation (Johnsson and Varshavsky, 1994b). Therefore, we compared the ratios of cleaved to uncleaved Suc2$_{23}$-C_{ub}-Dha and Mfα_{37}-C_{ub}-Dha. Cells expressing N_{ub}-Sec62p and either the C_{ub} fusion 8 or 10 (Figure 2B) were labeled for 5 min with ^{35}S-methionine and processed for immunoprecipitation with anti-ha antibody, followed by determination of the cleaved-to-uncleaved ratio (Figure 6D). This ratio, a measure of the time-averaged proximity of Sec62p to a translocating protein,

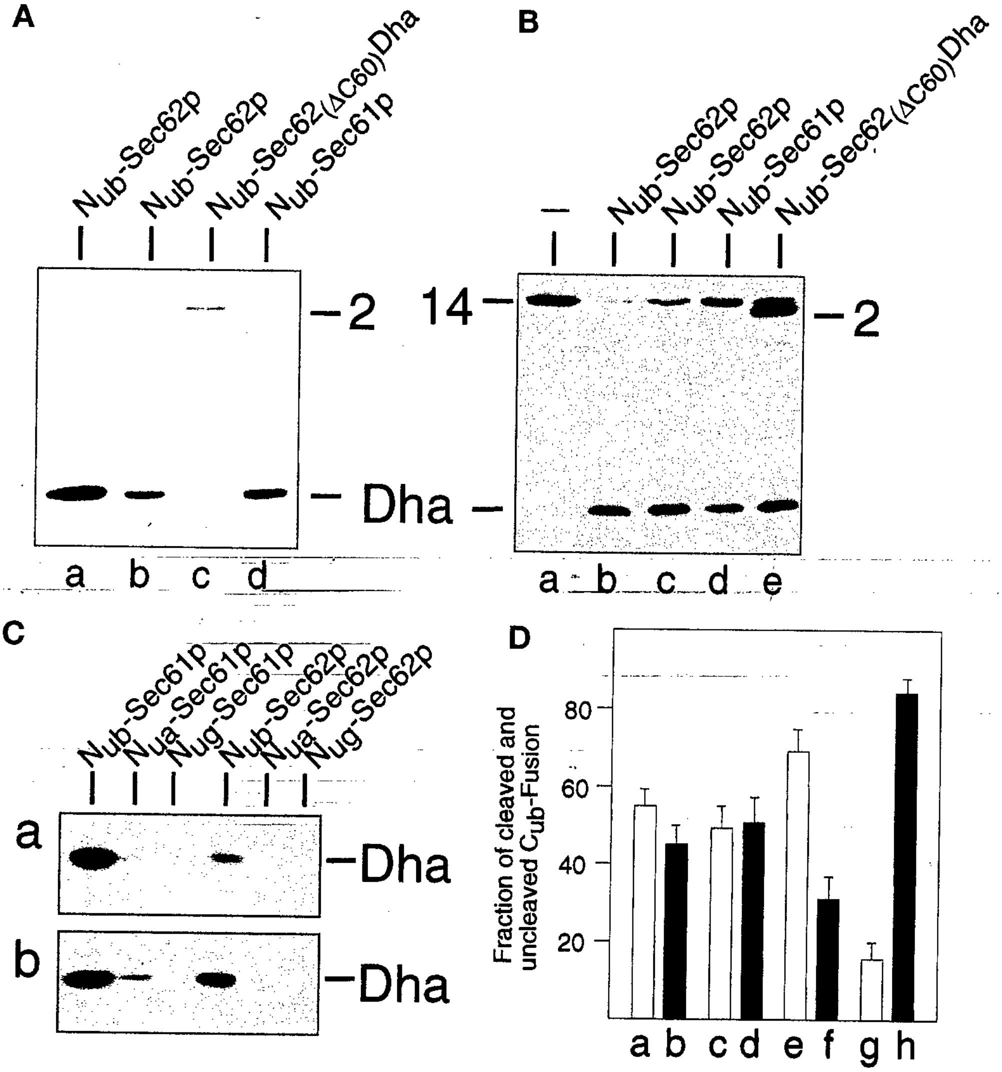


Figure 6. Sec61p, but not a mutant of Sec62p, are close to the nascent chain of a translocated protein. (A) These assays employed *S. cerevisiae* expressing Mfα_{37}-C_{ub}-Dha (construct 8, Figure 2) and one of the following N_{ub} fusions (Figure 2): N_{ub}-Sec62p, either integrated (lane a) or plasmid borne (lane b); N_{ub}-Sec62(ΔC60)Dha (lane c); and N_{ub}-Sec61p (lane d). Whole-cell extracts from these strains were subjected to immunoblot analysis with anti-ha antibody. (B) Same as panel A but the same N_{ub} fusions were coexpressed with Tpi1-C_{ub}-Dha (construct 14; Figure 2). Numbers 2 and 14 indicate the positions of the corresponding (uncleaved) fusions. (C) Lane a: *S. cerevisiae* expressing $Suc2_{23}$-C_{ub}-Dha (construct 10; Figure 2) together with either N_{ub}-Sec61p, N_{ua}-Sec61p, N_{ug}-Sec61p, N_{ub}-Sec62p, N_{ua}-Sec62p, or N_{ug}-Sec62p; lane b: same as lane a but cells expressed Mfα_{37}-C_{ub}-Dha (construct 8; Figure 2) instead of $Suc2_{23}$-C_{ub}-Dha. (D) *S. cerevisiae* cells expressing Mfα_{37}-C_{ub}-Dha together with N_{ub}-Sec61p (a and b) or N_{ub}-Sec62p (e and f), and cells expressing $Suc2_{23}$-C_{ub}-Dha together with N_{ub}-Sec61p (c and d) or N_{ub}-Sec62p (g and h) were labeled for 5 min with ^{35}S-methionine. Whole-cell extracts were immunoprecipitated with anti-ha antibody, followed by SDS-PAGE, and quantitation of the cleaved and uncleaved C_{ub} fusions using PhosphorImager. Shown are the relative amounts of the cleaved (white bars: a, c, e, and g) and uncleaved (black bars: b, d, f, and h) C_{ub} fusions. The sum of a cleaved and uncleaved fusion was set at 100 in each of the three independent experiments. SDs are also indicated.

was ~eightfold higher for a nascent polypeptide bearing the signal sequence of α-factor than for a nascent polypeptide bearing the invertase signal sequence (Figure 6D; compare Figures 4C and 5C).

Spacer sequences of different length or composition upstream of the C_{ub} moiety might nonspecifically influence the interaction between N_{ub} and C_{ub}. To assess this potential spacer effect, we constructed signal sequence-lacking versions of $Suc2_{23}$-C_{ub}-Dha and $Suc2_{59}$-C_{ub}-Dha (Figure 2B, C_{ub} fusions 15 and 16), and compared their ability to reconstitute Ub in the presence of coexpressed N_{ub}-Sec62p. Both of these C_{ub} fusions were cleaved at the C_{ub}-Dha junction at approximately the same rate in the presence of N_{ub}-Sec62p (Figure 5D), in contrast to the marked difference in the rate of cleavage observed for their signal sequence-bearing counterparts (Figure 5, B and C). This control experiment further emphasized the effect of distance between a signal sequence and the C_{ub} moiety on the efficiency of Ub reconstitution in the presence of N_{ub}-Sec62p. We conclude that the accessibility of Sec62p in vivo to a specific region of the nascent polypeptide chain is influenced by both the nature of a signal sequence and its distance from that region.

Sec61p Is Equidistant from Two Different Signal Sequences

A direct comparison between two different signal sequences upstream of the C_{ub} moiety presumes approximately equal residence times of the corresponding C_{ub} moieties in the cytosol. It is also essential to know that the influence of the identity of a signal sequence on the rate of Ub reconstitution is not due to a nonspecific intramolecular interaction. One way to address these issues involves measuring the reconstitution of Ub from the C_{ub} moieties of the fusions 8 and 10 (Figure 2B) and a N_{ub}-containing fusion that is not involved in translocation. As illustrated in Figure 4, this test is not feasible because of the rapid translocation of C_{ub}-containing constructs into the ER. Note, however, that since Sec61p is the central component of the translocation pore, proteins that utilize different targeting pathways will converge at Sec61p shortly before their translocation (Jungnickel and Rapoport, 1995). Taking advantage of this property of Sec61p, we assayed the proximity of Mfα_{37}-C_{ub}-Dha and Suc2_{23}-C_{ub}-Dha to N_{ub}-Sec61p. If Mfα_{37}-C_{ub}-Dha and Suc2_{23}-C_{ub}-Dha are cleaved at the C_{ub}–Dha junction equally well in the presence of N_{ub}-Sec61p, the above interpretation of the observed selectivity of N_{ub}-Sec62p toward Mfα_{37}-C_{ub}-Dha (Figures 5 and 6) would be confirmed.

To carry out this test, N_{ub} was fused to the cytosolic N terminus of Sec61p (Figure 2A) (Wilkinson *et al.*, 1996). N_{ub}-Sec61p is functionally active (Wittke and Johnsson, unpublished data). It induced the release of free Dha from of Mfα_{37}-C_{ub}-Dha and Tpi1p-Cub-Dha with efficiency similar to that of N_{ub}-Sec62p (Figure 6, A and B). Thus, the split-Ub assay independently confirmed that Sec61p is close to the nascent polypeptide chain during its translocation. To compare the in vivo interactions of Sec61p with the C_{ub} fusions 8 and 10, which bore different signal sequences (Figure 2B), the amount of free Dha was determined by immunoblotting of whole-cell extracts. It was found that in the presence of N_{ub}-Sec61p, similar amounts of Dha were released from Mfα_{37}-C_{ub}-Dha and Suc2_{23}-C_{ub}-Dha, whereas in the presence of N_{ub}-Sec62p twice as much Dha was released from Mfα_{37}-C_{ub}-Dha than from Suc2_{23}-C_{ub}-Dha (Figure 6C).

This result was confirmed and extended by labeling the cotransformed cells for 5 min with ^{35}S-methionine and quantifying the ratio of cleaved-to-uncleaved C_{ub} fusions (Figure 6D). As was already observed by the immunoblot analysis, the above ratio was ~1 for both Mfα_{37}-C_{ub}-Dha and Suc2_{23}-C_{ub}-Dha in the presence of N_{ub}-Sec61p, but ~2 for Mfα_{37}-C_{ub}-Dha, and ~0.25 for Suc2_{23}-C_{ub}-Dha in the presence of N_{ub}-Sec62p (Figure 6D). The difference revealed by the pulse-immunoprecipitation analysis is higher than the estimate obtained by the immunoblot analysis, most likely because of the continuous accumulation of cleaved (and long-lived) Dha before the processing of cells for immunoblotting.

A C-terminally Truncated Sec62p Is No Longer Proximal to the Signal Sequence

Does the proximity of Sec62p to a nascent polypeptide chain that is detected by the split-Ub assay reflect the physical binding of the signal sequence to this protein? We constructed a derivative of N_{ub}-Sec62p in which the C-terminal 60 residues of Sec62p were replaced by the DHFR-ha (Dha) moiety, yielding Sec62(ΔC60)-Dha. A similar Sec62p-invertase fusion was described by Deshaies and Schekman (1990) and shown to be nonfunctional. As expected, neither Sec62(ΔC60)-Dha nor N_{ub}-Sec62(ΔC60)-Dha complemented the ts phenotype of RSY529 cells (Figure 3B).

The Ub-reconstitution activity of N_{ub}-Sec62(ΔC60)-Dha in the presence of either Mfα_{37}-C_{ub}-Dha or Tpi1-C_{ub}-Dha (Figure 6, A and B) was compared with the activity of N_{ub}-Sec62p and N_{ub}-Sec61p in the presence of the same C_{ub}-containing fusions. Remarkably, no cleavage of Mfα_{37}-C_{ub}-Dha was observed in the presence of N_{ub}-Sec62(ΔC60)-Dha, whereas the cytosolic Tpi1-C_{ub}-Dha was cleaved. This result (Figure 6, A and B) indicated that the concentration and accessibility of N_{ub} were comparable for the functionally inactive N_{ub}-Sec62(ΔC60)-Dha and the functionally active N_{ub}-Sec62p. In these experiments, N_{ub}-Sec62(ΔC60)-Dha, which could be detected with the anti-ha antibody (Figure 6, A and B), was expressed from the uninduced P_{CUP1} promoter. Strikingly, even overexpression of N_{ub}-Sec62(ΔC60)-Dha, from the copper-induced P_{CUP1}, did not result in a significant cleavage of Mfα_{37}-C_{ub}-Dha (our unpublished results). These control experiments with the inactive derivative of Sec62p indicated that the proximity signal in the split-Ub assay with Sec62p requires the functional activity of Sec62p.

Using Ura3p Reporter to Detect the In Vivo Proximity between Sec62p and Signal Sequences

The DHFR-ha (Dha) reporter moiety of Mfα_{37}-C_{ub}-Dha was replaced by *S. cerevisiae* Ura3p (orotidine-5′-phosphate decarboxylase), yielding Mfα_{37}-C_{ub}-Ura3p. The use of cytosolic Ura3p as a reporter for translocation across membranes is well documented (Maarse *et al.*, 1992; Johnsson and Varshavsky, 1994b; Ng *et al.*, 1996). The high sensitivity of Ura3p-based assays (cells become Ura$^+$ if a threshold amount of Ura3p is present in the cytosol) allowed us to express the N_{ub} and C_{ub} fusions from the uninduced P_{CUP1} promoter. Since the efficient translocation of Mfα_{37}-C_{ub}-Ura3p sequesters the Ura3p activity in the ER, a *ura3Δ* strain of *S. cerevisiae* that expressed Mfα_{37}-C_{ub}-Ura3p remained Ura$^-$ (Figure 7A). N_{ub}-Sec62p, which, as shown above, is close to the nascent chain of Mfα_{37}-C_{ub}-Dha during its translocation, induced enough cleavage of Mfα_{37}-C_{ub}-Ura3p at the C_{ub}–Ura3p junction to render cells Ura$^+$ (Figure 7B). Cells were transformed with either N_{ub}-Sec62p, N_{ub}-Sec61p, N_{ub}-Sec62(ΔC60)-Dha, or N_{ub}-Bos1p to compare relative proximities of these N_{ub}-containing proteins to C_{ub} fusions bearing the Ura3p reporter moiety and either

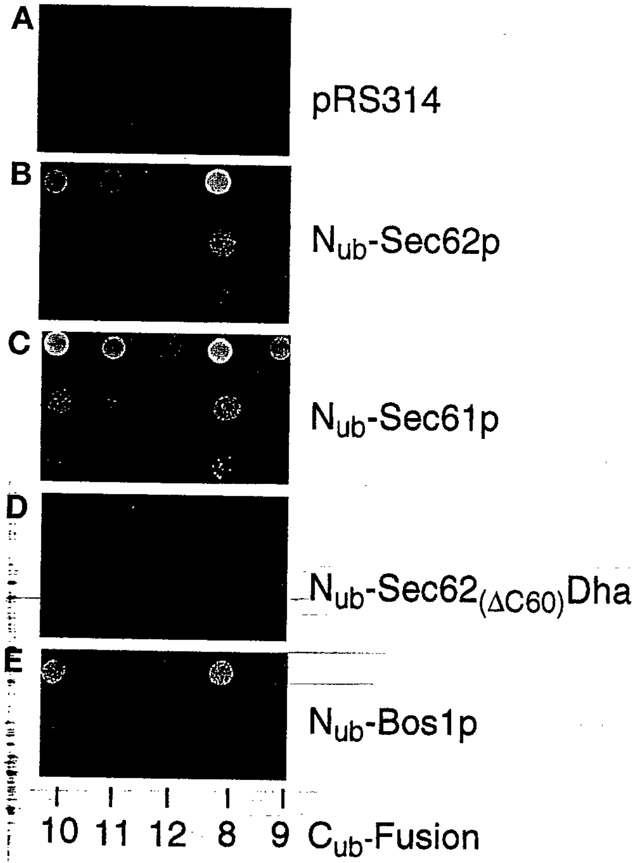

A
pRS314
B
N_{ub}-Sec62p
C
N_{ub}-Sec61p
D
N_{ub}-Sec62$_{(\Delta C60)}$Dha
E
N_{ub}-Bos1p
10 11 12 8 9 C_{ub}-Fusion

Figure 7. The use of a metabolic marker to assess the proximity between a component of the translocation machinery and a translocated protein. *S. cerevisiae* expressing the C_{ub} fusions 8–12 (Figure 2) that contained Ura3p instead of Dha (see the main text) were transformed with the vector (A) or plasmids expressing N_{ub}-Sec62p (B), N_{ub}-Sec61p (C), N_{ub}-Sec62(ΔC60)Dha (D), and N_{ub}-Bos1p (E). Cells were grown in a liquid uracil-containing SD medium, and ~10^5, 10^3, and 10^2 cells were spotted onto uracil-lacking SD medium. Plates were examined after 18 h at 30°C.

the invertase-derived or the α-factor-derived signal sequence (Figure 7, B–E). The cells were spotted on plates lacking uracil and incubated at 30°C for 18 h. The growth patterns of strains that expressed different combinations of N_{ub}- and C_{ub}-containing fusions confirmed the results of analyses with analogous (but more highly expressed) Dha-based constructs.

In particular, the interaction of Sec62p with the signal sequence of prepro-α-factor was stronger than with the signal sequence of invertase. This proximity was not detectable when the distance between a signal sequence and the C_{ub} moiety of a fusion was increased (Figure 7B). Sec61p appears to be equally close to both of the signal sequences tested. Again, the proximity signal was gradually lost when the distance between the signal sequence and the C_{ub} moiety was increased (Figure 7C). Cells acquired a weak Ura^+ phenotype in the presence of N_{ub}-Bos1p and the C_{ub}–Ura3p fusions 8 and 10 (Figure 7E). If used as a reference to discriminate between specific and nonspecific signals in this assay, Sec62p, under these conditions, appears to interact only with the α-factor signal sequence. No interaction with any of the tested C_{ub} constructs was detectable with the functionally inactive N_{ub}-Sec62(ΔC60)-Dha (Figure 7D).

DISCUSSION

The new application of the split-Ub technique (Johnsson and Varshavsky, 1994a, 1997) described in the present work introduces a tool for the analysis of transient (short-lived) protein interactions in living cells. A split-Ub assay involves the tagging of two (presumably) interacting proteins with the N- and C-terminal halves of Ub, N_{ub} and C_{ub}, and monitoring, in a variety of ways, the release of the reporter protein fused to the C terminus of C_{ub}. The reporter release, through the cleavage by Ub-specific processing proteases (UBPs), takes place in the cytosol if the two halves of Ub interact in vivo to form a quasi-native Ub moiety upstream of the reporter (Figure 1). Among the advantages of this method are its applicability either in living cells or in vitro and its sensitivity to kinetic aspects of a protein interaction.

In the present work, we applied the split-Ub technique to the problem of protein translocation across membranes. We showed that Sec62p of *S. cerevisiae* is spatially close to the signal sequence of the nascent α-factor polypeptide in vivo. This proximity is confined to the nascent polypeptide chain immediately following the signal sequence. In addition, the extent of proximity depends on the nature of the signal sequence. Specifically, C_{ub}-containing test proteins that bore the signal sequence of invertase resulted in a much lower Ub reconstitution with N_{ub}-Sec62p than the same C_{ub}-containing proteins bearing the signal sequence of α-factor. An inactive derivative of Sec62p failed to interact with signal sequences in the split-Ub assay. Taken together, these findings are the first in vivo evidence that *S. cerevisiae* Sec62p, a component of the ER translocation machinery, is a part of a signal sequence-binding complex.

In Vivo Proximity between Sec62p and the Signal Sequence of α-Factor

We have previously shown that a region of the nascent polypeptide chain that lies close to the signal sequence of invertase or the prepro-α-factor is briefly exposed to the cytosol before its translocation into the ER (Johnsson and Varshavsky, 1994b). This feature of translocation enabled us, in the present work, to use the split-Ub assay for monitoring the proximity between a secretory protein and components of the translocation machinery. The C_{ub} moiety was placed 37 residues downstream from the signal sequence of the α-factor precursor, and the N terminus of Sec62p (see INTRODUCTION) was extended with N_{ub}. Using this version of the split-Ub assay, we could demonstrate that Sec62p is close to the nascent chain of the α-factor during its translocation. By moving the C_{ub} moiety farther downstream from the signal sequence of the α-factor precursor, we obtained "snapshots" of the relative proximity between the nascent polypeptide chain and Sec62p in vivo. The proximity thus detected was considerably reduced once the spacer sequence between the signal sequence of the α-factor precursor and C_{ub} was increased from 37 to 65 residues (Figures 5B and 7B). The data strongly suggest that the access of Sec62p to the nascent chain of α-factor is confined to a region of the nascent chain that is very close to the signal sequence. Similar results were obtained with the signal sequence of invertase as well. This property of Sec62p is the expected feature of a component of a signal sequence receptor.

Our interpretation, supported by several control experiments, is in agreement with the results produced by cross-linking and binding studies in cell-free systems (Müsch *et al.*, 1992; Lyman and Schekman, 1997; Matlack *et al.*, 1997). Sec62p could be cross-linked to the α-factor precursor in vitro, but only when ATP was omitted and the initiation of translocation of α-factor was halted. Upon the addition of ATP, the translocation resumed and cross-linking was no longer possible (Müsch *et al.*, 1992; Lyman and Schekman, 1997). The cross-linking between Sec62p and the nascent polypeptide chain was not observed when the translocating chain was halted in the ER channel (Müsch *et al.*, 1992; Sanders *et al.*, 1992). It was therefore assumed that Sec62p is not a part of the channel and that it functions in the early steps of substrate recognition and initiation of translocation. The split-Ub assay, in its current form, depends on both halves of Ub being in the cell's cytosol. Therefore, the absence of the diagnostic cleavage (Figures 5B, 5C, and 7) when the C_{ub} moiety was placed farther downstream from the signal sequence (see RESULTS), while consistent with the absence of interactions between Sec62p and the nascent chain after the initiation of its translocation, does not address this issue directly. The C_{ub} moiety that emerges from the ribosome after it has docked at the ER channel is not accessible to the cytosolic N_{ub} moiety even if an N_{ub}-linked protein is spatially close to the translocation pore. This also explains the inability of N_{ub}-Sec61p to induce the cleavage of C_{ub}-containing translocation substrates bearing long spacer sequences between the signal sequence and C_{ub}, although the in vitro cross-linking studies have shown Sec61p to be in constant contact with the translocating polypeptide (Mothes *et al.*, 1994) (Figure 7).

The proximity between Sec62p (or Sec61p) and a translocating polypeptide is short lived. The rapid transfer of the C_{ub} moiety into the lumen of the ER was shown to either prevent or strongly inhibit its interaction with the N_{ub} moiety of N_{ub}-Sec62p and N_{ub}-Sec61p (Figures 3A and 6C). In these experiments, the N_{ub} moieties bore either the glycine (N_{ug}) or the alanine (N_{ua}) residue at position 13 of N_{ub}. These modifications decrease the affinity between the two halves of Ub (see INTRODUCTION), thereby making the reconstitution of a quasi-native Ub moiety more dependent on the stability (half-life) of interactions between the proteins linked to N_{ub} and C_{ub}. These assays clearly distinguished the Sec62p–Sec61p signal sequence interactions from those that underlie the better understood, longer-lived protein complexes. For example, when linked to homodimerizing leucine zippers, the N_{ua} moiety, and even the N_{ug} moiety, is sufficient for reconstitution of the Ub (Johnsson and Varshavsky, 1994a).

Is Sec62p Part of a Signal Sequence Receptor?

The split-Ub assay measures the concentrations of the protein-coupled N_{ub} and C_{ub} moieties in the immediate vicinity of each other. Therefore, a positive result of a split-Ub assay signifies a spatial proximity between the two proteins but cannot, by itself, prove their physical interaction or address the functional significance of this proximity. N_{ub}-Sec62(ΔC60)Dha is functionally inactive and was shown to be not close to the translocating $Mf\alpha_{37}$-C_{ub}-Dha (Figures 6 and 7). Since the sequence between N_{ub} and the first membrane-spanning region of Sec62 was retained in the C-terminally truncated Sec62(ΔC60)Dha, the distance between N_{ub} and the ER membrane was, most probably, not altered relative to wild-type Sec62p. The lack of significant cleavage of $Mf\alpha_{37}$-C_{ub}-Dha in the presence of N_{ub}-Sec62(ΔC60)Dha must therefore result from the increased distance between Sec62(ΔC60)p and the translocating polypeptide chain. The C-terminal domain of intact Sec62p may contact other components of the translocation complex; alternatively, it may contribute to a binding site for the signal sequence or the nascent chain. These and related uncertainties notwithstanding, our results (Figures 6 and 7) provide the first in vivo evidence to support the view that Sec62p is part of a signal sequence-binding complex.

Sec62p Discriminates between Different Signal Sequences

The split-Ub assay has made it possible to show that Sec62p discriminates, in living cells, between two distinct signal sequences in otherwise identical fusion proteins. In the presence of N_{ub}-Sec62p, more of the free Dha reporter protein was produced in vivo from $Mf\alpha_{37}$-C_{ub}-Dha than from $Suc2_{23}$-C_{ub}-Dha (Figures 5 and 6). This selectivity is a property of Sec62p and not a feature of the assay used, since approximately equal amounts of the cleaved reporter were produced with different signal sequences if the N_{ub} moiety was present as the N_{ub}–Sec61p fusion (Figures 6 and 7). This result confirmed, in vivo, that different targeting pathways converge at the Sec61-containing complex to initiate translocation. The existence of at least two different targeting pathways to the translocation pore was suggested by Walter and co-workers on the basis of the properties of yeast mutants that lacked either SRP or its receptor (Hann and Walter, 1991; Ogg *et al.*, 1992). One possibility is that the targeting via SRP operates cotranslationally, whereas the targeting via the Sec62/63 complex is predominantly posttranslational. The bifurcation between the cotranslational and posttranslational targeting is expected to be stochastic for many translocated proteins. Nonetheless, certain signal sequences do prefer SRP, while some of the other signal sequences are targeted by the Sec62/Sec63 complex (Ng *et al.*, 1996). Genetic studies have shown that the translocation of invertase continues in the presence of mutations in either the SRP or the Sec62/63 complex (Deshaies and Schekman, 1989; Ogg *et al.*, 1992). However, the kinetics and efficiency of invertase translocation are altered in the absence of SRP (Hann and Walter, 1991; Johnsson and Varshavsky, 1994b).

To explain the different efficiencies of cleavage of the two signal sequence-bearing C_{ub} fusions, we propose the following model. The signal sequence of invertase is recognized primarily by SRP and then transferred to the trimeric Sec61p complex, which completes the protein's translocation across the ER membrane. Under conditions that result in a shortage of SRP or a competition among different signal sequences, the Sec62p/Sec63p-containing complex, being a part of the alternative targeting pathway, would recognize an increasing fraction of invertase. This would explain why specific interactions between the invertase signal sequence-bearing proteins and Sec62p can only be observed for the more highly expressed C_{ub}–Dha fusions (Figure 5C). By contrast, the targeting of proteins bearing the signal sequence of the α-factor precursor is mediated, *in vivo*, predominantly by the Sec62p/Sec63p-containing complex. This would account for the observed close proximity of these test proteins to Sec62p (Figures 5–7). If so, the split-Ub technique makes it possible to estimate the flux of two different secretory proteins through the targeting pathways to the ER membrane without the necessity of deleting or otherwise inactivating specific components of the targeting complex.

Further Applications of the Split-Ub Technique

The split-Ub sensor should also be applicable to other settings that involve short-lived protein interactions that occur in the cytosol and are freely accessible to the Ub-specific proteases. The advantage of using this method for the analysis of protein translocation stems, in part, from the fast and irreversible removal of the translocated chain from the location (cytosol) where the N_{ub}/C_{ub} interaction is monitored. Similar situations are expected for the translocation of proteins into other organelles such as the mitochondrion, the nucleus, and the peroxisome.

The demonstration, in the present work, that Ura3p can serve as a reporter in a split-Ub assay (Figure 7) opens the way to genetic screens based on this assay. For example, introducing a DNA library consisting of random N_{ub}–gene fusions into a strain expressing a signal sequence-bearing C_{ub}-Ura3p protein should allow the identification of genes involved in targeting or translocation by enabling the cells to form colonies on media lacking uracil. This selection for protein ligands that interact transiently in the vicinity of a membrane complements the recent split-Ub–based screen for ligands that form relatively stable complexes (Stagljar *et al.*, 1998).

ACKNOWLEDGMENTS

We thank Ray Deshaies, Jürgen Dohmen, Nicole Lewke, Randy Schekman, and Sandra Wittke for the gifts of yeast strains and plasmids. M.D. and N.J. thank Silke Müller for excellent technical assistance. This work was supported by a grant to N.J. from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (0311107), and a grant to A.V. from the National Institutes of Health (GM-31530).

REFERENCES

Allison, D.S., and Young, E.T. (1988). Single-amino-acid substitutions within the signal sequence of yeast prepro-α-factor affect membrane translocation. Mol. Cell. Biol. *8*, 1915–1922.

Aronheim, A., Zandi, E., Hennemann, H., Elledge, S.J., and Karin, M. (1997). Isolation of an AP-1 repressor by a novel method for detecting protein- protein interactions. Mol. Cell. Biol. *17*, 3094–3102.

Banfield, D.K., Lewis, M.J., Rabouille, C., Warren, G., and Pelham, H.R. (1994). Localization of Sed5, a putative vesicle targeting molecule, to the *cis*- Golgi network involves both its transmembrane and cytoplasmic domains. J. Cell Biol. *127*, 357–371.

Beckmann, R., Bubeck, D., Grassucci, R., Penczek, P., Verschoor, A., Blobel, G., and Frank, J. (1997). Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. Science *278*, 2123–2126.

Biederer, T., Volkwein, C., and Sommer, T. (1996). Degradation of subunits of the Sec61p complex, an integral component of the ER membrane, by the ubiquitin-proteasome pathway. EMBO J. *15*, 2069–2076.

Bird, P., Gething, M.J., and Sambrook, J. (1987). Translocation in yeast and mammalian cells: not all signal sequences are functionally equivalent. J. Cell Biol. *105*, 2905–2914.

Brodsky, J.L., and Schekman, R. (1993). A Sec63p-BiP complex from yeast is required for protein translocation in a reconstituted proteoliposome. J. Cell Biol. *123*, 1355–1363.

Connolly, T., Collins, P., and Gilmore, R. (1989). Access of proteinase K to partially translocated nascent polypeptides in intact and detergent-solubilized membranes. J. Cell Biol. *108*, 299–307.

Crowley, K.S., Liao, S., Worrell, V.E., Reinhart, G.D., and Johnson, A.E. (1994). Secretory proteins move through the endoplasmic reticulum membrane via an aqueous, gated pore. Cell *78*, 461–471.

Crowley, K.S., Reinhart, G.D., and Johnson, A.E. (1993). The signal sequence moves through a ribosomal tunnel into a noncytoplasmic aqueous environment at the ER membrane early in translocation. Cell *73*, 1101–1115.

Deshaies, R.J., Sanders, S.L., Feldheim, D.A., and Schekman, R. (1991). Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. Nature *349*, 806–808.

Deshaies, R.J., and Schekman, R. (1989). SEC62 encodes a putative membrane protein required for protein translocation into the yeast endoplasmic reticulum. J. Cell Biol. *109*, 2653–2664.

Deshaies, R.J., and Schekman, R. (1990). Structural and functional dissection of Sec62p, a membrane-bound component of the yeast endoplasmic reticulum protein import machinery. Mol. Cell. Biol. *10*, 6024–6035.

Dohmen, R.J., Stappen, R., McGrath, J.P., Forrova, H., Kolarov, J., Goffeau, A., and Varshavsky, A. (1995). An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. J. Biol. Chem. *270*, 18099–18109.

Feldheim, D., and Schekman, R. (1994). Sec72p contributes to the selective recognition of signal peptides by the secretory polypeptide translocation complex. J. Cell Biol. *126*, 935–943.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245–246.

Gilmore, R., and Blobel, G. (1985). Translocation of secretory proteins across the microsomal membrane occurs through an environment accessible to aqueous perturbants. Cell *42*, 497–505.

Görlich, D., Prehn, S., Hartmann, E., Kalies, K.U., and Rapoport, T.A. (1992). A mammalian homolog of SEC61p and SECYp is associated with ribosomes and nascent polypeptides during translocation. Cell *71*, 489–503.

Hanein, D., Matlack, K.E., Jungnickel, B., Plath, K., Kalies, K.U., Miller, K.R., Rapoport, T.A., and Akey, C.W. (1996). Oligomeric rings of the Sec61p complex induced by ligands required for protein translocation. Cell *87*, 721–732.

Hann, B.C., and Walter, P. (1991). The signal recognition particle in *S. cerevisiae*. Cell *67*, 131–144.

Hiller, M.M., Finger, A., Schweiger, M., and Wolf, D.H. (1996). ER degradation of a misfolded luminal protein by the cytosolic ubiquitin-proteasome pathway. Science *273*, 1725–1728.

Johnsson, N., and Varshavsky, A. (1994a). Split ubiquitin as a sensor of protein interactions in vivo. Proc. Natl. Acad. Sci. USA *91*, 10340–10344.

Johnsson, N., and Varshavsky, A. (1994b). Ubiquitin-assisted dissection of protein transport across membranes. EMBO J. *13*, 2686–2698.

Johnsson, N., and Varshavsky, A. (1997). Split ubiquitin: a sensor of protein interactions in vivo. In: The Yeast Two Hybrid System, ed. P.L. Bartel and S. Fields, Oxford, United Kingdom: Oxford University Press, 316–332.

Jungnickel, B., and Rapoport, T.A. (1995). A posttargeting signal sequence recognition event in the endoplasmic reticulum membrane. Cell *82*, 261–270.

Krieg, U.C., Walter, P., and Johnson, A.E. (1986). Photocrosslinking of the signal sequence of nascent preprolactin to the 54-kilodalton polypeptide of the signal recognition particle. Proc. Natl. Acad. Sci. USA *83*, 8604–8608.

Kurzchalia, T.V., Wiedmann, M., Girshovich, A.S., Bochkareva, E.S., Bielka, H., and Rapoport, T.A. (1986). The signal sequence of nascent preprolactin interacts with the 54K polypeptide of the signal recognition particle. Nature *320*, 634–636.

Lyman, S.K., and Schekman, R. (1997). Binding of secretory precursor polypeptides to a translocon subcomplex is regulated by BiP. Cell *88*, 85–96.

Maarse, A.C., Blom, J., Grivell, L.A., and Meijer, M. (1992). MPI1, an essential gene encoding a mitochondrial membrane protein, is possibly involved in protein import into yeast mitochondria. EMBO J. *11*, 3619–3628.

Matlack, K.E., Plath, K., Misselwitz, B., and Rapoport, T.A. (1997). Protein transport by purified yeast Sec complex and Kar2p without membranes. Science *277*, 938–941.

Miyawaki, A., Llopis, J., Heim, R., McCaffery, J.M., Adams, J.A., Ikura, M., and Tsien, R.Y. (1997). Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. Nature *388*, 882–887.

Mothes, W., Prehn, S., and Rapoport, T.A. (1994). Systematic probing of the environment of a translocating secretory protein during translocation through the ER membrane. EMBO J. *13*, 3973–3982.

Müsch, A., Wiedmann, M., and Rapoport, T.A. (1992). Yeast Sec proteins interact with polypeptides traversing the endoplasmic reticulum membrane. Cell *69*, 343–352.

Nakano, A., Brada, D., and Schekman, R. (1988). A membrane glycoprotein, Sec12p, required for protein transport from the endoplasmic reticulum to the Golgi apparatus in yeast. J. Cell Biol. *107*, 851–863.

Ng, D.T., Brown, J.D., and Walter, P. (1996). Signal sequences specify the targeting route to the endoplasmic reticulum membrane. J. Cell Biol. *134*, 269–278.

Ogg, S.C., Poritz, M.A., and Walter, P. (1992). Signal recognition particle receptor is important for cell growth and protein secretion in *Saccharomyces cerevisiae*. Mol. Biol. Cell *3*, 895–911.

Orlean, P., Kuranda, M.J., and Albright, C.F. (1991). Analysis of glycoproteins from *Saccharomyces cerevisiae*. Methods Enzymol. *194*, 682–697.

Panzner, S., Dreier, L., Hartmann, E., Kostka, S., and Rapoport. T.A. (1995). Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. Cell *81*, 561–570.

Plemper, R.K., Böhmler, S., Bordallo, J., Sommer, T., and Wolf, D.H. (1997). Mutant analysis links the translocon and BiP to retrograde protein transport for ER degradation. Nature *388*, 891–895.

Rapoport, T.A., Jungnickel, B., and Kutay, U. (1996). Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. Annu. Rev. Biochem. *65*, 271–303.

Rossi, F., Charlton, C.A., and Blau, H.M. (1997). Monitoring protein-protein interactions in intact eukaryotic cells by β-galactosidase complementation. Proc. Natl. Acad. Sci. USA *94*, 8405–8410.

Rothblatt, J.A., Deshaies, R.J., Sanders, S.L., Daum, G., and Schekman, R. (1989). Multiple genes are required for proper insertion of secretory proteins into the endoplasmic reticulum in yeast. J. Cell Biol. *109*, 2641–2652.

Sanders, S.L., Whitfield, K.M., Vogel, J.P., Rose, M.D., and Schekman, R.W. (1992). Sec61p and BiP directly facilitate polypeptide translocation into the ER. Cell *69*, 353–365.

Sapperstein, S., Berkower, C., and Michaelis, S. (1994). Nucleotide sequence of the yeast STE14 gene, which encodes farnesylcysteine carboxyl methyltransferase, and demonstration of its essential role in a-factor export. Mol. Cell. Biol. *14*, 1438–1449.

Shim, J., Newman, A.P., and Ferro-Novick, S. (1991). The BOS1 gene encodes an essential 27-kDa putative membrane protein that is required for vesicular transport from the ER to the Golgi complex in yeast. J. Cell Biol. *113*, 55–64.

Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics *122*, 19–27.

Simon, S.M., and Blobel, G. (1991). A protein-conducting channel in the endoplasmic reticulum. Cell *65*, 371–380.

Stagljar, I., Korostensky, C., Johnsson, N., and te Heesen, S. (1998). A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. Proc. Natl. Acad. Sci. USA *95*, 5187–5192.

Walter, P., and Johnson, A.E. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. Annu. Rev. Cell Biol. *10*, 87–119.

Walter, P., Ibrahimi I., and Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum I. Signal recognition protein (SRP) binds to in vitro assembled polysomes synthesizing secretory protein. J. Cell Biol. *91*, 545–550.

Wiertz, E.J., Tortorella, D., Bogyo, M., Yu, J., Mothes, W., Jones, T.R., Rapoport, T.A., and Ploegh, H.L. (1996). Sec61-mediated transfer of a membrane protein from the endoplasmic reticulum to the proteasome for destruction. Nature *384*, 432–438.

Wilkinson, B.M., Critchley, A.J., and Stirling, C.J. (1996). Determination of the transmembrane topology of yeast Sec61p, an essential component of the endoplasmic reticulum translocation complex. J. Biol. Chem. *271*, 25590–25597.

The EMBO Journal Vol.18 No.21 pp.6017–6026, 1999

Degradation signals in the lysine–asparagine sequence space

Tetsuro Suzuki[1] and Alexander Varshavsky[2]

Division of Biology, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

[1]Present address: Department of Virology II, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan

[2]Corresponding author
e-mail: avarsh@its.caltech.edu

The N-degrons, a set of degradation signals recognized by the N-end rule pathway, comprise a protein's destabilizing N-terminal residue and an internal lysine residue. We show that the strength of an N-degron can be markedly increased, without loss of specificity, through the addition of lysine residues. A nearly exhaustive screen was carried out for N-degrons in the lysine (K)–asparagine (N) sequence space of the 14-residue peptides containing either K or N (16 384 different sequences). Of these sequences, 68 were found to function as N-degrons, and three of them were at least as active and specific as any of the previously known N-degrons. All 68 K/N-based N-degrons lacked the lysine at position 2, and all three of the strongest N-degrons contained lysines at positions 3 and 15. The results support a model of the targeting mechanism in which the binding of the E3–E2 complex to the substrate's destabilizing N-terminal residue is followed by a stochastic search for a sterically suitable lysine residue. Our strategy of screening a small library that encompasses the entire sequence space of two amino acids should be of use in many settings, including studies of protein targeting and folding.

Keywords: N-degron/N-end rule/proteolysis/simple sequences/ubiquitin

Introduction

Regulatory proteins are often short-lived *in vivo*, providing a way to generate their spatial gradients and to rapidly adjust their concentration or subunit composition through changes in the rate of their synthesis or degradation. Most of the damaged or otherwise abnormal proteins are metabolically unstable as well. Many other proteins, while long-lived as components of larger structures such as ribosomes and oligomeric proteins, are short-lived as free subunits (reviewed by Hochstrasser, 1996; Varshavsky, 1997; Hershko and Ciechanover, 1998; Scheffner *et al.*, 1998; Koepp *et al.*, 1999; Tyers and Willems, 1999).

Features of proteins that confer metabolic instability are called degradation signals, or degrons (Laney and Hochstrasser, 1999). One class of degradation signals, called the N-degrons, comprises a protein's destabilizing N-terminal residue and an internal Lys residue (Bachmair *et al.*, 1986; Varshavsky, 1996). A set of N-degrons containing different N-terminal residues that are destabilizing in a given cell defines a rule, termed the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. The lysine determinant of an N-degron is the site of formation of a substrate-linked multi-ubiquitin chain (Bachmair and Varshavsky, 1989; Chau *et al.*, 1989). The N-end rule pathway is thus one pathway of the ubiquitin (Ub) system. Ub is a 76-residue protein whose covalent conjugation to other proteins plays a role in a multitude of processes, including cell growth, division, differentiation, and responses to stress (Pickart, 1997; Varshavsky, 1997; Peters, 1998; Scheffner *et al.*, 1998). In many of these settings, Ub acts through routes that involve the degradation of Ub–protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease (Coux *et al.*, 1996; Hilt and Wolf, 1996; Baumeister *et al.*, 1998; Rechsteiner, 1998).

The N-end rule is organized hierarchically. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by the *NTA1*-encoded N-terminal amidase, into the secondary destabilizing N-terminal residues Asp and Glu (Baker and Varshavsky, 1995). The destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *ATE1*-encoded Arg-tRNA-protein transferase, to Arg, one of the primary destabilizing residues. The primary N-terminal residues are bound directly by the *UBR1*-encoded N-recognin, the E3 (recognition) component of the N-end rule pathway. In *S.cerevisiae*, N-recognin is a 225 kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues: Phe, Leu, Trp, Tyr, Ile, Arg, Lys and His (Varshavsky, 1996). Analogous components of the mammalian N-end rule pathway have been identified as well (Stewart *et al.*, 1995; Grigoryev *et al.*, 1996; Kwon *et al.*, 1998, 1999).

Studies with engineered N-end rule substrates indicated the bipartite organization of N-degrons and suggested a stochastic model of their targeting, in which specific lysines of an N-end rule substrate could be assigned different probabilities of being used as a ubiquitylation site (Bachmair and Varshavsky, 1989; Chau *et al.*, 1989; Johnson *et al.*, 1990; Hill *et al.*, 1993; Varshavsky, 1996; Lévy *et al.*, 1999). Most of the evidence for this model was produced with a set of N-degrons in which a destabilizing N-terminal residue X was linked to the ~40-residue *Escherichia coli* Lac repressor-derived sequence termed e^K [extension (e) bearing lysines (K)] (Figure 1A) (Bachmair and Varshavsky, 1989). The resulting X-e^K sequence comprised a portable N-degron, which could confer short half-lives on test proteins such as *E.coli* β-galactosidase

© European Molecular Biology Organization

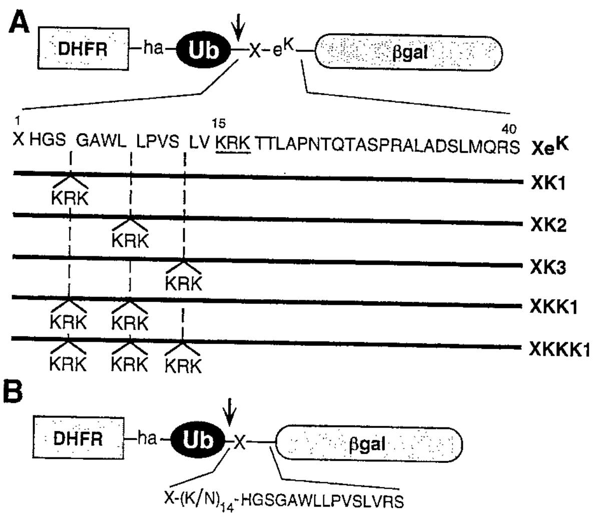


Fig. 1. Test proteins. (**A**) Fusions used in this work contained some of the following elements (see Materials and methods): DHFRha, a mouse dihydrofolate reductase moiety extended at the C-terminus by a sequence containing the hemagglutinin-derived ha epitope; the Ub^{R48} moiety bearing the Lys→Arg alteration at position 48; a 40-residue *E.coli* Lac repressor-derived sequence, termed e^K and shown in single-letter abbreviations for amino acids; a variable residue X (either Tyr, His or Met) between Ub^{R48} and e^K; the *E.coli* βgal moiety lacking the first 24 residues of wild-type βgal. The positions of single or multiple KRK insertions into e^K are indicated. The endogenous KRK sequence of e^K is underlined. The arrow indicates the site of *in vivo* cleavage by DUBs. (**B**) The fusion construct used for screening in the K/N sequence space. The 14 residues of e^K immediately following the residue X were replaced by a set of 14-residue sequences that comprised a random permutation of Lys and Asn residues, followed by the sequence HGSGAWLLPVSLVRS, derived from residues 2–14 of the e^K extension, followed by Arg-Ser.

(βgal) or mouse dihydrofolate reductase (DHFR) (Varshavsky, 1996). At least one of two lysines (K) in e^K, either K-15 or K-17, must be present for the N-degron to be active (Figure 1A) (Bachmair and Varshavsky, 1989; Johnson *et al.*, 1990). Even though several other classes of N-degron, including the naturally occurring ones, have been described over the last decade (Townsend *et al.*, 1988; deGroot *et al.*, 1991; Dohmen *et al.*, 1994; Sadis and Finley, 1995; Ghislain *et al.*, 1996; Sijts *et al.*, 1997; Tobery and Siliciano, 1999), the mechanistic understanding of these degradation signals remains confined largely to the e^K-based N-degrons (Varshavsky, 1996; Lévy *et al.*, 1999).

In the present work, we show that spiking an e^K-based N-degron with additional Lys residues can markedly increase its activity. We also show, using a new approach of searching in the sequence space of lysine and asparagine, that simple-sequence N-degrons can be as strong and specific as any of the previously known N-degrons. These findings provide independent evidence for the model of a bipartite N-degron and stochastic targeting mechanism (Bachmair and Varshavsky, 1989). The strategy of exhaustive searching in the sequence space of two amino acids should be of use in many settings, including studies of protein folding and degradation.

Results and discussion

The Ub/protein/reference technique

The assays below utilized the previously developed Ub/protein/reference (UPR) technique, which increases the accuracy of pulse–chase analysis by providing a 'built-in' reference protein (Lévy *et al.*, 1996). This method employs a linear fusion in which Ub is located between a protein of interest and a reference protein moiety (Figure 1A). The fusion is co-translationally cleaved by Ub-specific de-ubiquitylating enzymes (DUBs) (Wilkinson and Hochstrasser, 1998) after the last residue of Ub, producing equimolar amounts of the protein of interest and the reference protein bearing the C-terminal Ub moiety. If both the reference protein and the protein of interest are immunoprecipitated in a pulse–chase assay, the relative amounts of the protein of interest can be normalized against the reference protein in the same sample. The UPR technique can thus compensate for the scatter of immunoprecipitation yields, sample volumes and other sources of sample-to-sample variation (Lévy *et al.*, 1996, 1999).

Two previously introduced terms, ID^x, initial decay, i.e. the decay of a protein during the pulse of x min, and $t_{0.5}^{y-z}$, the protein's half-life averaged over the interval of y to z min of chase (Lévy *et al.*, 1996), are used below to describe the decay curves of test proteins. The ID^x term and the interval-specific term $t_{0.5}^{y-z}$ would be superfluous in the case of a strictly first-order decay, which is defined by a single half-life. However, the *in vivo* degradation of most proteins deviates from first-order kinetics. For example, the rate of degradation of short-lived proteins can be much higher during the pulse, in part because a newly labeled (either nascent or just-completed) polypeptide is conformationally immature and may, consequently, be targeted for degradation more efficiently than its mature counterpart. This enhanced early degradation, previously termed the 'zero-point' effect (Baker and Varshavsky, 1991), is described by the parameter ID^x (Lévy *et al.*, 1996). It was found that a large fraction of the zero-point effect results from the co-translational degradation of nascent (being synthesized) polypeptide chains, which never reach their mature size before their destruction by processive proteolysis (G.Turner and A.Varshavsky, unpublished data). The detection of a zero-point effect requires the comparison of a test protein's degradation between cells containing and lacking the relevant proteolytic pathway. Alternatively, the zero-point effect can be detected by comparing, through the UPR technique, the degradation of otherwise identical degron-containing and degron-lacking versions of a test protein (Lévy *et al.*, 1996, 1999). Although the degradation of a protein during the pulse can be strikingly high (Lévy *et al.*, 1996) (see also below), it is not detectable by a conventional, reference-lacking pulse–chase assay.

Increasing the strength of N-degrons by spiking them with additional lysines

The UPR constructs of the present work were DHFR-ha-Ub^{R48}-X-e^K-βgal fusions. They contained the metabolically stable, ha-epitope-bearing DHFR-ha-Ub^{R48} moiety as a reference protein, termed dha-Ub below. The dha-Ub-X-e^K-βgal proteins were co-translationally cleaved *in vivo*, yielding the test protein X-e^K-βgal and the reference dha-Ub (Figure 1A). To reduce the possibility that the C-terminal Ub moiety of dha-Ub could function as a ubiquitylation/degradation signal, the K-48 residue of Ub (a major site of isopeptide bonds in multi-Ub chains) was converted to Arg, which cannot be ubiquitylated, yielding Ub^{R48} (Lévy *et al.*, 1996). These and related fusions (Figure 1A) were expressed in *S.cerevisiae* from low copy plasmids and the copper-inducible P_{CUP1} promoter.

X-e^{K}-βgal is an extensively analyzed class of N-end rule substrates, which contain a variable N-terminal residue X (produced through the DUB-mediated cleavage of dha-Ub-X-e^{K}-βgal at the Ub–X junction), a 40-residue N-terminal extension called e^{K} (see Introduction), and a βgal moiety lacking the first 24 residues of wild-type *E.coli* βgal (Figure 1A). If K-15 and K-17, the only lysines of the e^{K} extension (Figure 1A), are replaced by Arg residues, which cannot be ubiquitylated, the resulting X-e$^{\Delta K}$-βgal is long-lived even if its N-terminal residue is destabilizing in the N-end rule (Bachmair *et al.*, 1986; Johnson *et al.*, 1990). The inactivity of N-degron in X-e$^{\Delta K}$-βgal is caused by the absence of targetable lysines (Varshavsky, 1996). Specifically, the multiple lysines of the βgal moiety in X-e$^{\Delta K}$-βgal (Chau *et al.*, 1989) cannot serve as N-degron determinants, apparently because the most N-terminal Lys residue in X-e$^{\Delta K}$-βgal, at position 239, is too far from the protein's N-terminus.

One of our aims was to produce stronger N-degrons. We chose Tyr, a moderately destabilizing type 2 residue (Bachmair and Varshavsky, 1989; Varshavsky, 1996), as the N-terminal residue of an initial test protein (Figure 1A). More strongly destabilizing N-terminal residues, e.g. Leu or Arg, in the context of (expected) stronger N-degrons would have made the test proteins too short-lived for detection in a pulse–chase assay. Met was employed as a stabilizing N-terminal residue. The term ID^{5} below (see Materials and methods) conveys the extent of degradation of a protein during the 5 min pulse, in comparison with the degradation, during the same pulse, of a control (degron-lacking, i.e. Met-bearing) version of the same protein.

To determine whether the degradation of Tyr-e^{K}-βgal in *S.cerevisiae* could be enhanced through the addition of Lys residues while remaining dependent on the Ubr1p N-recognin, the sequence Lys-Arg-Lys (KRK), identical to the sequence at positions 15–17 of e^{K}, was inserted at the indicated locations within e^{K} (Figure 1A). The unmodified Tyr-e^{K}-βgal had an ID^{5} of ~48%, i.e. ~48% of the labeled Tyr-e^{K}-βgal was destroyed during the 5 min pulse, before time 0. The $t_{0.5}^{0-10}$ (half-life between 0 and 10 min of chase) of Tyr-e^{K}-βgal was ~26 min (Figures 2A and 3A). The KRK sequence inserted at any of the indicated three positions within e^{K} (Figure 1A) strongly destabilized the already short-lived Tyr-e^{K}-βgal: for example, Tyr-K1e^{K}-βgal (Figure 1A) had an ID^{5} of ~75% and $t_{0.5}^{0-10}$ of ~5 min (Figures 2A and 3A). The increased degradation of Tyr-e^{K}-βgal derivatives containing extra KRK remained completely Ubr1p-dependent: Tyr-e^{K}-βgal, Tyr-K1e^{K}-βgal, Tyr-K2e^{K}-βgal and Tyr-K3e^{K}-βgal were all long-lived proteins ($t_{0.5}$ >10 h) in *ubr1*Δ cells (Figures 2A and 3A). In addition, Met-e^{K}-βgal, Met-K1e^{K}-βgal, Met-K2e^{K}-βgal and Met-K3e^{K}-βgal, the Met-bearing counterparts of the Tyr-based N-end rule substrates, were long-lived in either *UBR1 or ubr1*Δ cells (data not shown).

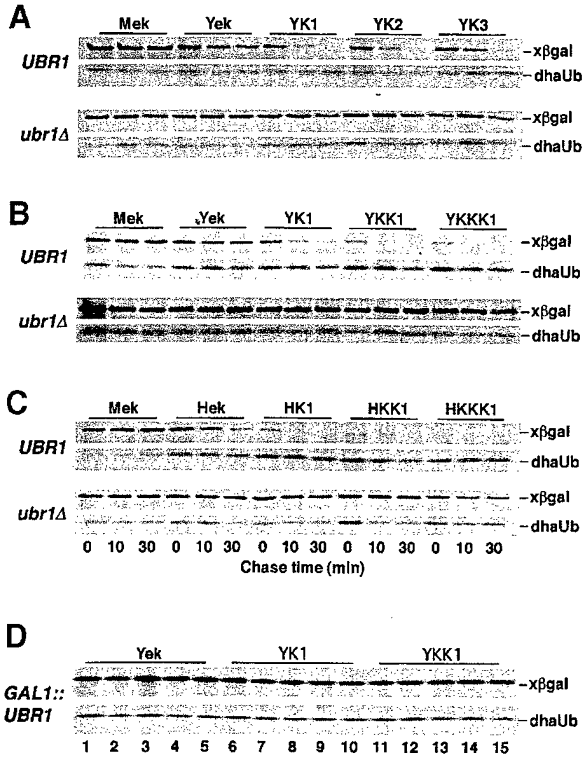


Fig. 2. Active N-degrons can be strongly enhanced by additional lysines. (**A**) Congenic *UBR1* (wt) and *ubr1*Δ *S.cerevisiae* that expressed the UPR-based fusions Met-e^{K}-βgal (Mek) (DHFR-ha-Ub^{R48}-Met-e^{K}-βgal), Tyr-e^{K}-βgal (Yek), Tyr-K1e^{K}-βgal (YK1), Tyr-K2e^{K}-βgal (YK2) and Tyr-K3e^{K}-βgal (YK3) (see Figure 1A) were labeled with [^{35}S]methionine/cysteine for 5 min at 30°C, followed by a chase for 0, 10 and 30 min, extraction, immunoprecipitation with anti-ha and anti-βgal antibodies, SDS–PAGE, and autoradiography (see Materials and methods). The bands of X-βgal (test protein) and DHFR-ha-Ub^{R48} (reference protein) are indicated on the right. (**B**) As in (A), but with Met-e^{K}-βgal (Mek), Tyr-e^{K}-βgal (Yek), Tyr-K1e^{K}-βgal (YK1), Tyr-KK1e^{K}-βgal (YKK1) and Tyr-KKK1e^{K}-βgal (YKKK1) (see Figure 1A). (**C**) As in (A), but with Met-e^{K}-βgal (Mek), His-e^{K}-βgal (Hek), His-K1e^{K}-βgal (HK1), His-KK1e^{K}-βgal (HKK1) and His-KKK1e^{K}-βgal (HKKK1). (**D**) Metabolic stability of conformationally mature Tyr-e^{K}-βgal and its KRK-spiked derivatives. JD54 (P_{GAL1}-*UBR1*) cells expressing Tyr-e^{K}-βgal (Yek), Tyr-K1e^{K}-βgal (YK1) or Tyr-KK1e^{K}-βgal (YKK1) were grown in SM-raffinose medium (no expression of Ubr1p), then labeled with [^{35}S]methionine/cysteine for 10 min at 30°C. After a 20 min chase in SM-raffinose, galactose was added to 3% to induce Ubr1p expression, followed by a chase for 1, 3 and 6 h, and the analysis of immunoprecipitated test proteins. Lanes 1, 6 and 11, the end of ^{35}S labeling (time 0). Lanes 2, 7 and 12, the end of 20 min chase in SM-raffinose. Lanes 3, 8 and 13, 1 h chase with galactose. Lanes 4, 9 and 14, 3 h chase. Lanes 5, 10 and 15, 6 h chase.

These results led us to examine the effects of adding more than one KRK sequence to e^{K} (Figure 1A). The resulting Tyr-e^{K}-βgal derivatives, bearing either two (Tyr-KK1e^{K}-βgal) or three (Tyr-KKK1e^{K}-βgal) KRK sequences, in addition to the original KRK of e^{K}, were extremely short-lived proteins, even though N-terminal Tyr is a weakly destabilizing residue (Varshavsky, 1996). For example, Tyr-KK1e^{K}-βgal (Figure 1A) had $t_{0.5}^{0-10}$ of ~4 min (in comparison with ~26 min in the case of Tyr-e^{K}-βgal) and an ID^{5} of ~94%. In other words, ~94% of the labeled Tyr-KK1e^{K}-βgal was destroyed during the 5 min pulse, before time 0 (Figures 2B and 3B). At the same time, all of these proteins were long-lived in *ubr1*Δ cells (Figures 2B and 3B).

The N-terminal Tyr is bound by the type 2 site of N-recognin (Ubr1p) that recognizes substrates bearing bulky hydrophobic N-terminal residues (Varshavsky, 1996). We asked whether the above findings were also relevant to the type 1 (basic) destabilizing N-terminal residues Arg, Lys and His, which are bound by the type 1 site of Ubr1p. Counterparts of the Tyr-βgal fusions that bore N-terminal His (a weak type 1 destabilizing residue) were constructed (Figure 1A) and tested in pulse–chase assays. The His residue was chosen as a type 1 destabilizing residue in these tests for the same reason as Tyr in the preceding tests: a stronger destabilizing residue would have made the measurements impractical with extremely short-lived substrates. The results (Figures 2C and 3C) confirmed the generality and specificity of degradation enhancement by the additional KRK sequences. For example, His-K1e^{K}-βgal had an ID^{5} of ~95% (i.e. ~95% of the labeled His-K1e^{K}-βgal was destroyed during the 5 min pulse, before time 0), in comparison with the ID^{5} of ~41% for His-e^{K}-βgal; the corresponding $t_{0.5}^{0\text{-}10}$ values were ~5 min and ~8 min for His-K1e^{K}-βgal and His-e^{K}-βgal, respectively (Figure 3C).

Similar to the results with Tyr-bearing substrates, their His-bearing, multiple KRK-containing counterparts were long-lived in *ubr1Δ* cells (Figures 2C and 3C). The only exception was His-KKK1e^{K}-βgal (Figure 1A), which contained three KRK sequences, in addition to the KRK of the original e^{K}: in contrast to Tyr-KKK1e^{K}-βgal, His-KKK1e^{K}-βgal was stabilized strongly but incompletely in the *ubr1Δ* genetic background (Figure 3C). Thus, the His-KKK1e^{K} extension, in contrast to the Tyr-KKK1e^{K} extension (Figure 1A), appears to contain a Ubr1p-independent degron.

Previous work (Madura *et al.*, 1993; Kwon *et al.*, 1999) has shown that the steady-state level of an X-βgal protein (determined by measuring the enzymatic activity of βgal in yeast extracts) is a sensitive measure of its metabolic stability. The results of this steady-state assay were in agreement with those derived from pulse–chase measurements: the level of Tyr-e^{K}-βgal in *UBR1* cells was 53% of the level of the long-lived Met-e^{K}-βgal, whereas Tyr-K1e^{K}-βgal was present at 4% of the Met-e^{K}-βgal level, and the concentration of Tyr-KKK1e^{K}-βgal was virtually indistinguishable from the assay's background (cells transformed with vector alone) (Figure 3E). Crucially, the levels of these extra KRK-bearing Tyr-e^{K}-βgal fusions in *ubr1Δ* cells became similar to that of Met-e^{K}-βgal (Figure 3E), in agreement with the pulse–chase data (Figures 2A, B and 3A, B).

Although the addition of extra KRK sequences to e^{K} yielded considerable decreases in the $t_{0.5}^{0\text{-}10}$ of the corresponding N-end rule substrates, by far the major effect of multiple KRK sequences was on the decay curves' ID^{5} term, which conveys the extent of degradation of a protein during or shortly after its synthesis (Figures 2A–C and 3A–C). To examine this issue in a different way, Tyr-e^{K}-βgal, Tyr-K1e^{K}-βgal and Tyr-KK1e^{K}-βgal (Figure 1A) were produced in the JD54 *S.cerevisiae* strain, which expressed Ubr1p from the galactose-inducible, dextrose-repressible P$_{GAL1}$ promoter. JD54 cells expressing one of the test proteins were labeled in raffinose-containing SR medium (no Ubr1p), incubated for 20 min in the same medium and thereafter shifted to galactose, where Ubr1p was induced. Even though the N-end rule pathway became hyperactive in the presence of galactose (Madura and Varshavsky, 1994; Ghislain

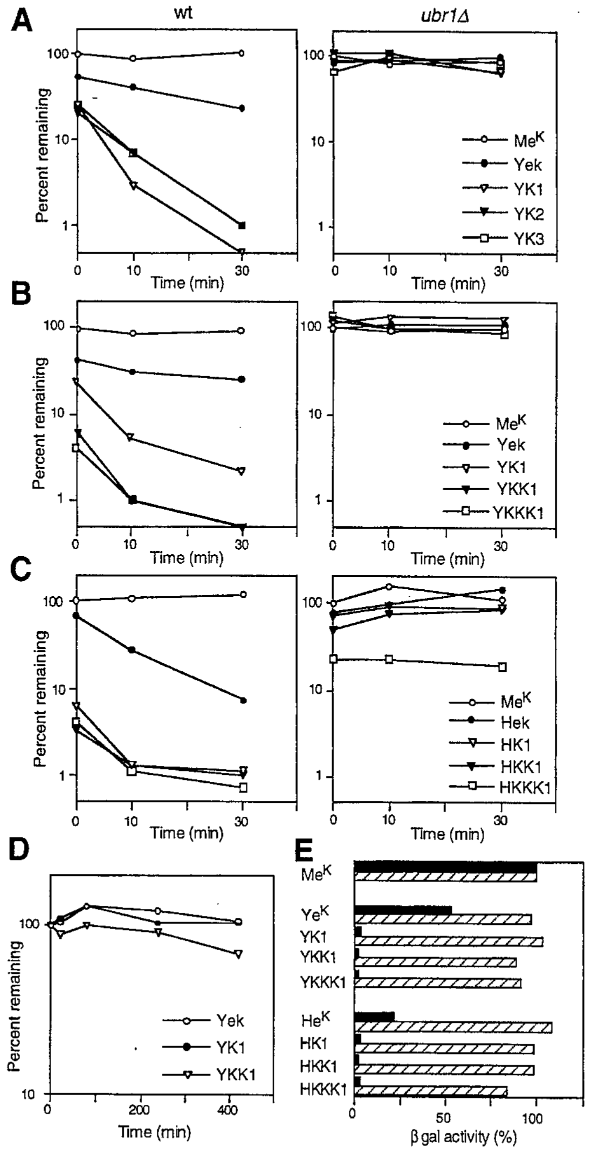


Fig. 3. Quantitation of degradation of the test proteins. (**A**) Pulse–chase patterns of Met-e^{K}-βgal (Me^{K}), Tyr-e^{K}-βgal (Ye^{K}), Tyr-K1e^{K}-βgal (YK1), Tyr-K2e^{K}-βgal (YK2) and Tyr-K3e^{K}-βgal (YK3) (see Figures 1A and 2A) were quantitated using the UPR technique and PhosphorImager (see Materials and methods). Time 0 refers to the end of the 5 min pulse; 100% refers to the relative amount of Met-e^{K}-βgal, normalized against the reference protein dha-Ub^{R48}. (**B**) As in (A) but with Met-e^{K}-βgal (Me^{K}), Tyr-e^{K}-βgal (Ye^{K}), Tyr-K1e^{K}-βgal (YK1), Tyr-KK1e^{K}-βgal (YKK1) and Tyr-KKK1e^{K}-βgal (YKKK1) (see Figures 1A and 2B). (**C**) As in (A) but with Met-e^{K}-βgal (Me^{K}), His-e^{K}-βgal (He^{K}), His-K1e^{K}-βgal (HK1), His-KK1e^{K}-βgal (HKK1) and His-KKK1e^{K}-βgal (HKKK1) (see Figures 1A and 2C). (**D**) As in (A), but quantitation of the post-translational degradation of Tyr-e^{K}-βgal (Ye^{K}), Tyr-K1e^{K}-βgal (YK1) and Tyr-KK1e^{K}-βgal (YKK1) in the P$_{GAL1}$-*UBR1* strain JD54 (see Figures 1A and 2A), following the induction of Ubr1p by galactose. Time 0 refers to the end of the 10 min labeling in raffinose (no Ubr1p). The cells were incubated for another 20 min in raffinose, followed by the addition of galactose to induce Ubr1p (see Materials and methods). (**E**) Relative enzymatic activities of βgal in *UBR1* cells (filled bars) and *ubr1Δ* cells (striped bars) expressing one of the following test proteins: Met-e^{K}-βgal (Me^{K}), Tyr-e^{K}-βgal (Ye^{K}), Tyr-K1e^{K}-βgal (YK1), Tyr-KK1e^{K}-βgal (YKK1), Tyr-KKK1ue^{K}-βgal (YKKK1), His-e^{K}-βgal (He^{K}), His-K1e^{K}-βgal (HK1), His-KK1e^{K}-βgal (HKK1) and His-KKK3e^{K}-βgal (HKKK1) (see Materials and methods). The activities of βgal were normalized to the activity of Met-e^{K}-βgal in each cell. Values shown are the means from duplicate measurements, which yielded results within 10% of the mean values.

et al., 1996; data not shown), the pre-labeled substrates Tyr-e^K-βgal and Tyr-K1e^K-βgal were barely degraded after the induction of Ubr1p; Tyr-KK1e^K-βgal was degraded only slightly (Figures 2D and 3D).

These findings were consistent with the earlier evidence for a strong retardation of the post-translational degradation of Arg-e^K-βgal under the same conditions (R.J.Dohmen and A.Varshavsky, unpublished data). Thus, in contrast to a newly formed, conformationally immature βgal-based test protein, a conformationally mature βgal tetramer is a poor substrate of the N-end rule pathway even in the presence of N-degron enhancements such as the additional KRK sequences. It is the βgal moiety of these test proteins (Figure 1A) that was responsible for the time-dependent decline in the rate of degradation, because the kinetics of *in vivo* degradation of e^K-DHFR-based N-end rule substrates was much closer to first-order decay (Lévy *et al.*, 1999; data not shown).

Locating N-degrons in the lysine–asparagine sequence space

The earlier work, which led to the bipartite model of N-degron (Bachmair and Varshavsky, 1989; Hill *et al.*, 1993), and particularly the present findings about the effects of adding KRK sequences to an e^K-based N-degron (Figures 2 and 3) suggested that a substrate's destabilizing N-terminal residue and a sterically suitable internal Lys residue (or residues) are the two necessary and sufficient components of an N-degron. However, since both the e^K-based and other previously analyzed N-degrons are embedded in complex sequence contexts (deGroot *et al.*, 1991; Dohmen *et al.*, 1994; Varshavsky, 1996), we wished to address the bipartite-degron model by constructing an N-degron from much simpler sequence motifs. Should this prove feasible, we also wanted to explore constraints on the structure of N-degrons through a screen in a simpler sequence setting. If the sequence space could be reduced strongly enough, one advantage of such a screen would be its exhaustiveness. The AAA codon for lysine differs by just one third-letter substitution from the codon for asparagine (AAU), a polar uncharged residue. Thus, one could attempt a screen for N-degrons in the sequence space of two amino acids: Lys (K) and Asn (N).

A double-stranded oligonucleotide that encoded random 14-residue K/N sequences (see Materials and methods) was used to replace the sequence encoding 14 residues of e^K immediately following the residue X (Figure 1B). In the resulting test proteins, this latter sequence, HGSGAWLLPVSLVRS (plus the sequence RS), followed the quasi-random 14-residue K/N sequence (Figure 1A). The resulting K/N-based extensions either lacked the lysines or contained a variable number of them between residues 2 and 16. The K-17 of e^K (the only other lysine in e^K) was replaced by Arg. In these test fusions, dha-Ub^{R48}-Arg-(K/N)$_{14}$-e$^\Delta$-βgals, Arg was used as a destabilizing N-terminal residue (Figure 1B). The number of different 14-residue sequences containing exclusively K or N is 2^{14} = 16 384. The bulk of a library of this complexity could be encompassed with conventional screening methods. Testing of the pRKN14-based library by amplifying it in *E.coli* indicated that >90% of the plasmids contained an oligonucleotide insert. The pRKN14 library was introduced into *S.cerevisiae* JD54 (Ghislain *et al.*, 1996), which expressed Ubr1p from the P$_{GAL1}$ promoter, and screened for colonies that stained blue with XGal [high levels of Arg-(K/N)$_{14}$-e$^\Delta$-βgal] on dextrose (SD) plates but stained white [low levels of Arg-(K/N)$_{14}$-e$^\Delta$-βgal] on replica-plated galactose (SG) plates. Approximately 20 000 colonies were screened this way. A total of 68 isolates were identified in which the activity of βgal was significantly higher in the absence than in the presence of the N-end rule pathway.

The corresponding Arg-(K/N)$_{14}$-e$^\Delta$-βgal test proteins were expressed in congenic *ubr1Δ* and *UBR1* strains, and the ratio of βgal activities was determined for each of the test proteins. The results are summarized in Figure 4, which shows the K/N sequences of the 30 most active N-degrons, and the ratios of the corresponding βgal activity in the *ubr1Δ* strain to that in the *UBR1* strain (higher ratios indicate stronger N-degrons). Remarkably, the strongest K/N-based N-degron was found to be more active than the strongest e^K-based N-degron (Figure 4). Black bars in Figure 4 denote βgal activity derived from constructs carrying K/N-based N-degrons with lysines present at positions 3 and 15; the strongest N-degrons were largely of this class (Figure 4). K-15 was present in the 15 strongest K/N-based N-degrons except one (clone 132), which had K at position 14, and was also, presumably in compensation for the absence of K-15, one of the most lysine-rich N-degrons in this set (Figure 4). Similarly, K-3 was present in the 15 strongest K/N-based N-degrons except three (clones 3, 77 and 138) (Figure 4). All of these exceptional clones bore K-15; in addition, one of them (clone 77) bore K-4 and K-5, as well as K-14 and K-15 (Figure 4).

A completely uniform feature of all 68 K/N-based N-degrons was the absence of K from position 2 (Figure 4; data not shown), consistent with the fact that all of the previously examined N-degrons (Varshavsky, 1996) also lacked a strongly basic residue (Arg or Lys) at position 2. To address this issue directly, we used site-directed mutagenesis, replacing asparagine at position 2 of clone 119 (the strongest K/N-based N-degron; Figure 4) with lysine. The resulting test protein was long-lived in *UBR1* cells (data not shown), confirming that lysine is not tolerated at position 2 of an N-degron.

Even though all of the strongest K/N-based N-degrons contained at least three lysines, the total number of lysines in a K/N-based N-degron was not a strong predictor of its activity, and the arrangement of additional lysines between positions 3 and 15 did not correlate, in an obvious way, with the activity of N-degrons (Figure 4). None of the 15 strongest K/N-based N-degrons had more than six lysines (most had from three to five) (Figure 4), indicating that the density of lysines *per se* is not the main feature of a strong K/N-based N-degron. Note that our screen rejected K/N-based degrons that exhibited a significant Ubr1p-independent activity. This may in part account for the upper limit on the number of lysines in the strongest N-degrons: more lysine-rich K/N sequences could contain motifs recognized by non-N-end rule pathways, similarly to the above His-KKK1-βgal test protein (Figures 1A, 2C and 3C).

The *in vivo* degradation of Arg-(K/N)$_{14}$-e$^\Delta$-βgal proteins bearing K/N-based N-degrons [produced from the dha-Ub^{R48}-Arg-(K/N)$_{14}$-e$^\Delta$-βgal fusions (Figure 1B)] was com-

Residue position

Clone	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Re^{k}	H	G	S	G	A	W	L	L	P	V	S	L	V	K
119	-	K	-	-	-	-	K	-	-	-	-	-	-	K
113	-	K	-	-	-	-	-	-	K	-	-	-	-	K
96	-	K	-	-	-	-	-	-	-	K	-	-	-	K
123	-	K	-	-	-	-	-	-	-	K	-	-	K	K
118	-	K	-	-	-	-	-	-	-	-	-	K	-	K
26	-	K	-	-	-	-	-	-	-	K	-	K	-	K
3	-	-	-	-	K	-	-	-	-	-	-	-	K	K
74	-	K	-	-	K	-	-	K	-	-	K	-	-	K
69	-	K	-	-	-	-	-	-	-	-	K	-	-	K
130	-	K	-	K	-	-	-	-	-	K	-	K	-	K
77	-	-	K	K	-	-	-	-	-	-	-	-	K	K
138	-	-	-	-	-	-	-	K	-	K	K	K	-	K
132	-	K	-	-	-	K	K	-	K	K	-	-	K	-
19	-	K	-	-	-	K	-	-	-	-	-	-	-	K
75	-	K	-	-	-	-	-	-	-	-	-	K	-	K
87	-	K	-	-	K	K	-	-	-	-	-	K	K	-
20	-	K	K	K	-	-	-	K	-	-	-	-	-	K
7	-	-	-	-	-	-	-	K	K	-	-	K	-	K
65	-	-	-	-	K	-	K	-	-	-	K	K	-	K
76	-	-	-	K	-	K	K	-	-	-	K	-	K	K
94	-	-	-	-	-	K	-	K	K	-	-	-	-	K
105	-	-	-	-	K	-	K	-	-	-	K	-	-	K
67	-	-	-	K	K	-	K	K	-	-	-	-	-	K
129	-	-	-	-	-	-	K	-	-	-	-	K	K	K
56	-	-	-	-	K	-	K	-	-	K	-	-	-	K
103	-	-	-	-	-	-	K	-	-	K	-	-	-	K
98	-	-	-	K	K	K	-	-	-	-	-	-	-	K
142	-	K	-	-	-	-	-	K	K	-	-	-	-	-
124	-	-	-	-	-	-	K	-	K	-	-	-	-	K
59	-	-	-	K	-	-	K	-	-	K	-	-	-	K

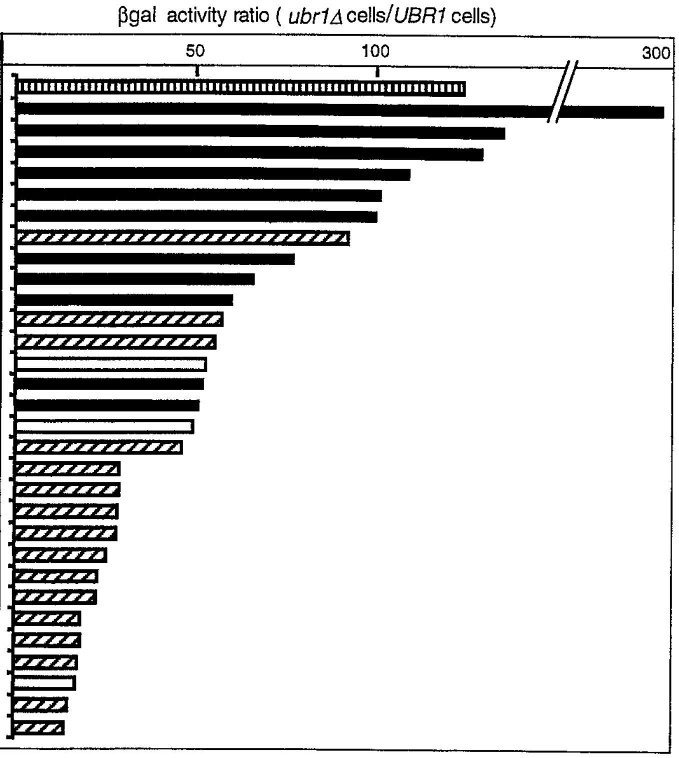


Fig. 4. N-degrons in the K/N sequence space. The deduced sequences of the identified K/N N-degrons are shown in conjunction with the bar diagram of their relative activity, defined as the ratio of βgal activities in the *ubr1Δ* versus *UBR1* cells expressing a given Arg-(K/N)$_{14}$-e^{Δ}-βgal test protein. The top bar (vertical stripes) indicates the relative activity of N-degron in the original Arg-e^{K}-βgal (Re^{K}). Thirty 14-residue K/N extensions with the highest Ubr1p-dependent destabilizing activity are listed, out of the total of 68 isolates that were metabolically unstable in the presence but not in the absence of Ubr1p. The total number of possible 14-residue K/N sequences is 16 384 (see the main text). The Lys and Asn residues are denoted as the letter K and a hyphen, respectively. Position 1 in each clone was occupied by Arg. The extensions in which lysines were present either at positions 3 and 15, or only at 15, or only at 3, are marked, respectively, by the filled, striped and open bars.

pared with the degradation of Arg-e^{K}-βgal in pulse–chase assays (Figure 5A and B). The chosen K/N degrons (clones 119, 113 and 4) were the two most active ones in the βgal assay (clones 119 and 113) (Figure 4), and a relatively weak one (clone 4) (Figure 6A). As expected from the results of steady-state βgal assays (Figures 4), Arg-(K/N)$_{14}$-e^{Δ}-βgal proteins 119 and 113 were extremely short-lived in *UBR1* cells (Figure 5A), in contrast to their stability in *ubr1Δ* cells (Figure 5B). As with Arg-e^{K}-βgal, the bulk of degradation of Arg-(K/N)$_{14}$-e^{Δ}-βgals took place either during or shortly after their synthesis, so that even at time 0 (at the end of the 5 min pulse) the test proteins could be detected only by overexposing the autoradiograms (Figure 5A; data not shown; compare with Figure 5B). In contrast, Arg-(K/N)$_{14}$-e^{Δ}-βgal bearing the weaker N-degron of clone 4 was readily detectable at time 0, and decayed more slowly afterwards (Figure 5A), in agreement with the results of steady-state βgal assays (Figure 6B). The distribution of three lysines in the strongest N-degron (clone 119: Lys3, 8, 15) (Figure 4) was similar to that in a relatively weak one (clone 4: Lys4, 12, 14) (Figure 6), emphasizing the importance of lysines at positions 3 and 15.

To address in more detail the relative contributions of the lysines at positions 3 and 15 to the activity of a K/N-based N-degron, site-directed mutagenesis was used to construct the otherwise identical N-degrons that contained either exclusively Lys3 and Lys15 (no lysine at a third, interior position), exclusively Lys3 or exclusively Lys15 (Figure 6A). The resulting Arg-(K/N)$_{14}$-e^{Δ}-βgal proteins were examined using both steady-state (Figure 6B) and pulse–chase assays (Figure 5C and D). The K/N-based N-degron that contained only Lys3 and Lys15 was active but considerably less so than the strongest K/N-based N-degron of clone 119, which contained Lys8 as well (Figures 4, 5C and 6). The elimination of either Lys3 or Lys15 further weakened the N-degron, so that the Lys3 and Lys15 versions of Arg-(K/N)$_{14}$-e^{Δ}-βgal were, respectively, slightly and moderately short-lived proteins (Figures 5C, D and 7).

Mechanistic implications

One finding of this work is that an N-degron can be greatly strengthened by spiking it with additional lysine residues. The resulting degradation signals remained specific: nearly all of the enhanced N-degrons were completely inactive in *ubr1Δ* cells, which lacked the E3 (recognition) component of the N-end rule pathway. In addition, the enhanced N-degrons could be inactivated by replacing their destabilizing N-terminal residue with Met, a stabilizing residue in the N-end rule.

The strength of a K/N-based N-degron depends on the arrangement of lysines in the 14-residue N-terminal region of the test protein. This dependence is both strong and complex (Figures 4 and 6). The patterns observed may reflect distinct conformational flexibilities of different K/N sequences *vis-à-vis* the relatively fixed spatial arrangement of the type 1 or 2 sites of Ubr1p (N-recognin) and its associated Ubc2p E2 enzyme (Madura *et al.*, 1993).

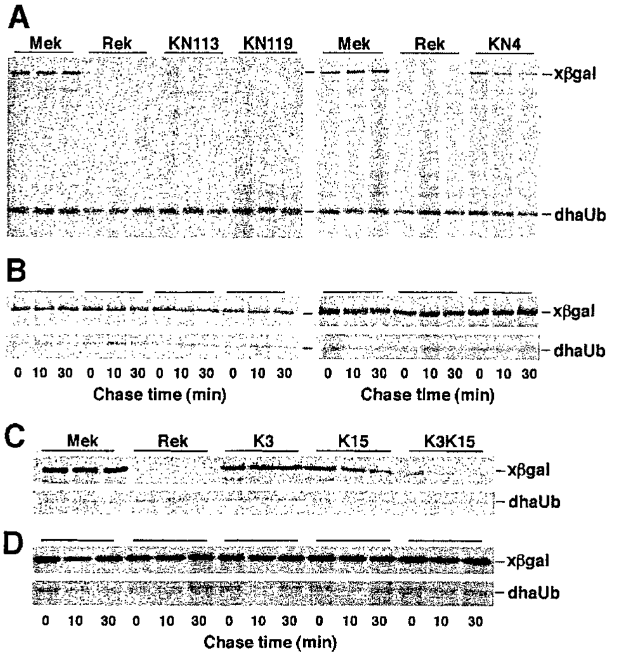


Fig. 5. Pulse–chase analysis of K/N-based N-degrons. (**A**) JD47-13C (*UBR1*) *S.cerevisiae* that expressed the UPR-based fusions Met-e^K-βgal (Mek) (DHFR-ha-Ub^{R48}-Met-e^K-βgal), R-e^K-βgal (Rek), clone 113 of Arg-(K/N)$_{14}$-e$^{\Delta}$-βgal (KN113), clone 119 of Arg-(K/N)$_{14}$-e$^{\Delta}$-βgal (KN119), clone 4 of Arg-(K/N)$_{14}$-e$^{\Delta}$-βgal (KN4) and other clones (see Figures 1B and 4) were labeled with [^{35}S]methionine/cysteine for 5 min at 30°C, followed by a chase for 0, 10 and 30 min, extraction, immunoprecipitation with anti-ha and anti-βgal antibodies, SDS–PAGE and autoradiography (see Materials and methods). The bands of X-βgal (test protein) and DHFR-ha-Ub^{R48} (reference protein) are indicated on the right. (**B**) As in (A), but with the congenic JD55 (*ubr1*Δ) cells. (**C**) As in (A) but with the clones K3, K15 and K3K15 (see Figure 6). (**D**) As in (C) but with JD55 (*ubr1*Δ) cells.

Some of the conclusions indicated by our data are described below.

(i) A completely uniform feature of all 68 K/N-based N-degrons was the absence of Lys from position 2 (Figure 4; data not shown), consistent with the fact that all of the previously examined N-degrons (Varshavsky, 1996) also lacked a strongly basic residue (Arg or Lys) at position 2. Using site-directed mutagenesis with clone 119 (Figure 4), we confirmed that lysine is not tolerated at position 2 of an N-degron (data not shown).

(ii) The strongest K/N-based N-degrons contained single Lys residues, surrounded by Asn residues (Figure 4), in contrast to the Lys-Arg-Lys (KRK) motif present in the e^K extension of N-degrons studied previously (Figure 1A). We conclude that a lysine-containing motif of three adjacent basic residues is not an essential feature of an N-degron.

(iii) The presence of Lys15 in all of the most active K/N-based N-degrons (Figure 4) is consistent with the earlier model (Bachmair and Varshavsky, 1989; Varshavsky, 1996) in which a targetable Lys residue should be sufficiently far along the chain from a destabilizing N-terminal residue to allow the formation of a loop that positions this lysine spatially close to the N-terminal residue. This model is also consistent with the finding that Lys3 is another, nearly invariant, component of a K/N-based N-degron. Specifically, Lys3 may be located at the uniquely favorable 'linear' distance from the N-terminal residue, allowing its proximity to the destabilizing N-terminal residue in the absence of loop formation. That this view is at best incomplete is indicated by the finding that either Lys3 alone or Lys15 alone cannot substitute for the combination of Lys3, Lys8 and Lys15 that defines the strongest K/N-based N-degron identified to date (clone 119 in Figures 4 and 6). Moreover, even Lys3 and Lys15 together, in the absence of Lys8, result in a much weaker N-degron that the three-lysine N-degron of clone 119 (Figure 6). One possibility is that Lys15 is the only lysine, among the three, that can function as the site of ubiquitylation, and that lysines at positions 3 and 8 are required largely for optimal conformational flexibility

A

Clone	Residue position														βgal activity ratio: *ubr1Δ* cells/*UBR1* cells
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
K3	-	K	-	-	-	-	-	-	-	-	-	-	-	-	1.5
K15	-	-	-	-	-	-	-	-	-	-	-	-	-	K	1.9
K3K15	-	K	-	-	-	-	-	-	-	-	-	-	-	K	11.8
119	-	K	-	-	-	-	K	-	-	-	-	-	-	K	299.0
113	-	K	-	-	-	-	-	-	K	-	-	-	-	K	135.2
4	-	-	K	-	-	-	-	-	-	-	K	-	K	-	7.3

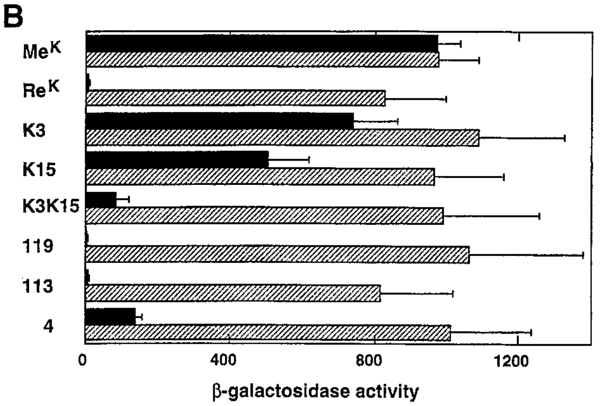


Fig. 6. Site-directed mutagenesis of N-degrons identified through the screen in the lysine–asparagine sequence space. (**A**) K/N sequences of clones 119, 113 and 4 and the K3/K15 derivatives of clone 119 (K3, K15 and K3K15). The relative activity of these N-degrons, defined as the ratio of βgal activities in the *ubr1*Δ versus *UBR1* cells expressing a given Arg-(K/N)$_{14}$-e$^{\Delta}$-βgal protein, is indicated on the right. (**B**) The levels of βgal activity in extracts from *UBR1* cells (filled bars) and *ubr1*Δ cells (striped bars) cells expressing the indicated test proteins. Standard deviations (for triplicate measurements) are indicated.

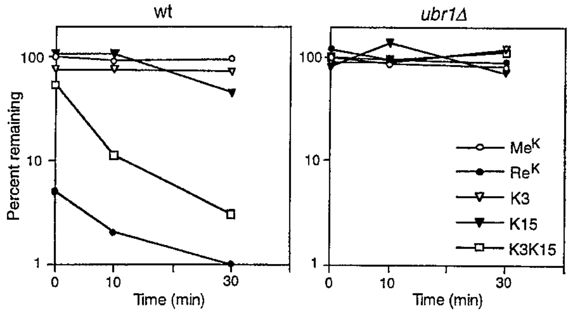


Fig. 7. Decay curves of K3 and/or K15 mutants of Arg-(K/N)$_{14}$-e$^{\Delta}$-βgal in *UBR1 and ubr1*Δ *S.cerevisiae*. Pulse–chase patterns of Met-e^K-βgal (Me^K), Arg-e^K-βgal (Re^K), and Lys3 (K3), Lys15 (K15) and Lys3/Lys15 (K3K15) variants of Arg-(K/N)$_{14}$-e$^{\Delta}$-βgal in Figure 5C and D were quantitated as described in the legend to Figure 3 and Materials and methods.

of the 14-residue K/N region in this N-degron. It is unlikely that the much higher activity of the clone 119 (three-lysine) N-degron resulted simply from its higher positive charge, in comparison with the two-lysine N-degrons, because most of the other K/N sequences containing three or more lysines were much less active than the clone 119 N-degron (Figure 4).

We searched for N-degrons in a K/N region 14 residues long (Figures 1B and 4), in part because the search with a significantly longer K/N region would have precluded the screen from being exhaustive. Nevertheless, since the e^{K} extension of the earlier N-degrons is 40 residues long (Figure 1A), and since a part of the e^{K} extension was retained, downstream of the K/N region, in our test proteins (see Materials and methods), it remains to be determined whether a K/N-based N-degron could be made even stronger by replacing the entire e^{K} with a relevant K/N motif. For example, one could take advantage of the already identified 14-residue K/N sequence of clone 119 (Figure 4), and carry out an analogous screen for N-degrons in which the K/N sequence of clone 119 is fixed, while the downstream, e^{K}-derived sequence is replaced by random K/N motifs. In sum, the K/N strategy of this work is far from exhausted by the present screen, and can be used, for example, to produce even stronger N-degrons, to search for other K/N-based degradation signals and to probe the targeting mechanisms of the Ub system.

Concluding remarks

One implication of our results is the possibility of constructing much stronger N-degrons. These portable degradation signals can be used to render proteins of interest that are short-lived in either conditional or unconditional settings (Dohmen *et al.*, 1994; Worley *et al.*, 1998).

Another implication of our results stems from the demonstrated feasibility of defining a specific class of degradation signals in a sequence space of just two amino acids. Since the N-degrons have been defined previously in much more complex sequence contexts (Varshavsky, 1996), our findings suggest that other classes of degradation signal, and some of the other targeting signals as well, could also be identified and examined in this simple-sequence setting. Low-complexity sequences eliminate some of the informational 'noise' of natural sequences and thereby help to define the major determinants of structural specificity. The advantages of the simple-sequence approach have long been recognized by researchers who study the fundamentals of protein folding (Clarke, 1995). To our knowledge, the present work is the first to extend the simple-sequence approach to the realm of protein degradation.

Materials and methods

Strains, media and genetic techniques

Saccharomyces cerevisiae strains used in this work were JD47-13C (*MAT***a** *ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3 112*) (Madura *et al.*, 1993), JD55 (*MAT***a** *ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3 112 ubr1-Δ1::HIS3*) (Madura and Varshavsky, 1994) and JD54 (*MAT***a** *ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3 112 GAL1::UBR1*) (Ghislain *et al.*, 1996). Rich (YPD) medium contained 1% yeast extract, 2% peptone (Difco) and 2% glucose. Synthetic media (Ausubel *et al.*, 1996) contained either 2% dextrose (glucose) (SD medium), 3% raffinose (SR medium) or 3% galactose (SG medium). To induce the P$_{CUP1}$ promoter, CuSO$_4$ was added to a final concentration of 0.2 mM. Cells were incubated and assayed at 30°C. Transformation of *S.cerevisiae* was carried out using the lithium acetate method (Ausubel *et al.*, 1996).

Plasmid construction

The plasmids encoded Ub fusions of the UPR technique (Lévy *et al.*, 1996)] (see Results and discussion). The reference protein was mouse DHFR fused, through a 20-residue spacer containing the ha epitope, to the Ub^{R48} moiety bearing Arg instead of wild-type Lys at position 48 (Lévy *et al.*, 1996). The plasmids pDhaUbXeKβgal expressed, from the P$_{CUP1}$ promoter, the fusions DHFR-ha-Ub^{R48}-X-e^{K}-βgal, where the junctional residue X was either Met, Tyr or His; e^{K} was a previously described 40-residue, *E.coli* Lac repressor-derived N-terminal extension (Bachmair and Varshavsky, 1989; Johnson *et al.*, 1992); the *E.coli* βgal moiety lacked the first 24 residues of wild-type βgal (Bachmair *et al.*, 1986). The βgal-coding fragment was produced by PCR amplification of the βgal open reading frame (ORF) in pUB23 (Bachmair *et al.*, 1986; Bachmair and Varshavsky, 1989), using primers BGALS1 (5′-CAGAGATCTCTTAATCGCCTTGCAGCA-3′) and RSAS2 (5′-CCCTCGAGGTCGACGGTATCG-3′), and the Expand PCR System (Boehringer, Indianapolis, IN). The *Bgl*II/*Xho*I-cut PCR product was ligated to the insert-lacking, P$_{CUP1}$ promoter-containing *Bgl*II–*Xho*I fragment of pdUbXekUra3, a pRS314-based, low copy plasmid (F.Lévy and A.Varshavsky, unpublished data). ORFs that expressed DHFR-ha-Ub^{R48}-X-e^{K}-βgal containing a modified e^{K} region [single, double or triple insertions of the sequence Lys-Arg-Lys (KRK) (Figure 1A)] were produced using PCR. Specifically, pDhaUbXK1βgal, pDhaUbXK2βgal, pDhaUbXK3βgal, pDhaUbXKK1βgal and pDhaUbXKKK1βgal were constructed by replacing the *Bam*HI–*Bgl*II fragment encoding X-e^{K}-βgal with DNA fragments amplified using, respectively, the sense primers EKKS1 (5′-CACGGATCCAAGAGAAAGGGAGCTTGGCTGTTGCCC-3′), EKKS2 (5′-CACGGATCCGGAGCTTGGCTGAAGAGAAAGTTGCCCGTCTCACTGGTG-3′), EKKS3 (5′-CACGGATCCGGAGCTTGGCTGTTGCCCGTCTCAAAGAGAAAGCTGGTGAAAAGAAAAACC-3′), EKKS4 (5′-CACGGATCCAAGAGAAAGGGAGCTTGGCTGAAGAGAAAGTTGCCCGTCTCACTGGTG-3′) and EKKKS1 (5′-CACGGATCCAAGAGAAAGGGAGCTTGGCTGAAGAGAAAGTTGCCCGTCTCAAAGAGAAAGCTGGTGAAAAGAAAAACC-3′). The primer EKAS1 (5′-GGAAGATCTCTGCATTAATGAATC-3′) and the plasmid pDhaUbMeKβgal served as the antisense primer and the template, respectively. The plasmids pDhaUbRK3eΔβgal, pDhaUbRK15eΔβgal and pDhaUbRK3K15eΔβgal, which expressed, respectively, DHFR-ha-Ub^{R48}-Arg-Asn-Lys-(Asn)$_{12}$-e$^{\Delta}$-βgal, DHFR-ha-Ub^{R48}-Arg-(Asn)$_{13}$-Lys-e$^{\Delta}$-βgal and DHFR-ha-Ub^{R48}-Arg-Asn-Lys(Asn)$_{11}$-Lys-e$^{\Delta}$-βgal, were constructed as follows. The term e$^{\Delta}$ denotes the sequence HGSGAWLLPVSLVRS, a 13-residue derivative of the 40-residue e^{K} (residues 2–14), followed by two residues Arg and Ser, encoded by the *Bgl*II site (Figure 1B). Double-stranded oligonucleotides that encompassed the 5′ end of the *Sac*II site, the end of Ub-Arg-(Lys/Asn)$_{14}$-coding sequences (see below), and the 3′ end of the *Bam*HI site were digested with *Sac*II and *Bam*HI. The resulting fragments were ligated to the *Sac*II–*Bam*HI vector-containing fragment of pDhaUbReΔβgal. The latter plasmid was produced by inserting annealed complementary oligonucleotides EDKS1 and EDAS1, which encoded the e$^{\Delta}$ sequence HGSGAWLLPVSLVRS (see above), into *Bam*HI/*Bgl*II-cut pDhaUbReKβgal. All constructs were verified by nucleotide sequencing.

A screen for N-degrons in the lysine–asparagine sequence space

The library of pDhaUbReΔβgal-derived plasmids was constructed that expressed, from the P$_{CUP1}$ promoter, Arg-(K/N)$_{14}$-e$^{\Delta}$-βgal proteins that contained random-sequence 14-residue Lys/Asn(K/N)-inserts between the N-terminal Arg and the e$^{\Delta}$ moiety (Figure 1B). Arg-(K/N)$_{14}$-e$^{\Delta}$-βgals were the products of co-translational cleavage of DHFR-ha-Ub^{R48}-Arg-(K/N)$_{14}$-e$^{\Delta}$-βgals (Figure 1B). To produce the library, a method for cloning random-sequence oligonucleotides produced by mutually primed synthesis was used (Oliphant *et al.*, 1986; Ghislain *et al.*, 1996). A set of oligonucleotides 5′-CGCCCGCGGTGGTAGG(AAA/T)$_{14}$CACGGATCCG-3′ that contained random permutations of 14 codons, either AAA (encoding Lys) or AAT (encoding Asn), flanked by the *Sac*II and *Bam*HI sites (underlined), was synthesized. The oligonucleotides were converted into their double-stranded counterparts with Klenow Pol I (Ausubel *et al.*, 1996), and were digested with *Sac*II and *Bam*HI, yielding a set of equal-length fragments containing randomly permuted (AAA/T)$_{14}$ inserts. The fragments were ligated to the *Sac*II–*Bam*HI vector-containing fragment of pDhaUbReΔβgal. The resulting library, in the plasmid termed pRKN14, was introduced into *E.coli* DH5α by electroporation (Ausubel *et al.*, 1996). Digestion of the pool of recovered plasmids with appropriate restriction enzymes showed that >90% of transformants contained an oligonucleotide-derived insert. The pRKN14 library was transformed into *S.cerevisiae* JD54, in which Ubr1p was expressed from the P_{GAL1} promoter. Approximately 2×10^4 transformants growing on SD(–Trp) plates containing 0.2 mM $CuSO_4$ were replica-plated onto $CuSO_4$-containing SG(–Trp) plates. The XGal-based filter assay (Ausubel *et al.*, 1996) was used to screen for colonies that were white (low levels of βgal) on SG(–Trp) plates but blue (high levels of βgal) on SD(–Trp) plates. These colonies were grown up in liquid SG(–Trp), and the activity of βgal was determined as described below. Plasmid DNA was isolated from the positive transformants, amplified in *E.coli* and transformed into *S.cerevisiae* JD47-13C (*UBR1*) and JD55 (*ubr1Δ*). The metabolic stabilities of the corresponding Arg-$(K/N)_{14}$-e^{Δ}-βgal proteins in *ubr1Δ* versus wild-type (*UBR1*) *S.cerevisiae* were determined by measuring βgal activity, and also directly, by carrying out pulse–chase assays.

Measurement of βgal activity

Saccharomyces cerevisiae cells were added to 5 ml of SD(–Trp) containing 0.2 mM $CuSO_4$, and grown to A_{600} ~1. Cells were gently pelleted by centrifugation, lysed with glass beads in 20% glycerol, 1 mM dithiothreitol, 0.1 M Tris–HCl pH 8, and the activity of βgal was measured in the clarified extract using *o*-nitrophenyl-β-D-galactoside, as described (Ausubel *et al.*, 1996). The activity was normalized to the total protein concentration, determined using the Bradford assay (Bio-Rad, Hercules, CA).

Pulse–chase assays

Transformed JD47-13C and JD55 cells from 10 ml cultures (A_{600} 0.5–1) in SD(–Trp) containing 0.2 mM $CuSO_4$ were gently pelleted by centrifugation, and washed in the same medium. The cells were resuspended in 0.3 ml of SD(–Trp) containing $CuSO_4$ and labeled for 5 min at 30°C with 0.15 mCi (5.5 MBq) of [^{35}S]methionine/cysteine (Trans^{35}S-label, ICN, Costa Mesa, CA). The cells were harvested by centrifugation, resuspended in 0.3 ml of SD(–Trp), 10 mM L-methionine, 0.5 mg/ml cycloheximide and incubated further at 30°C. At each time point, 0.1 ml samples were withdrawn and added to 0.7 ml of the lysis buffer (1% Triton X-100, 0.15 M NaCl, 1 mM EDTA, 50 mM Na-HEPES pH 7.5) containing 1 mM phenylmethylsulfonyl fluoride. The cells were then lysed by vortexing with 0.5 ml of 0.5 mm glass beads four times for 1 min, with intermittent cooling on ice, followed by centrifugation at 12 000 *g* for 10 min. The volumes of supernatants were adjusted to equalize the amounts of 10% trichloroacetic acid-insoluble ^{35}S, followed by immunoprecipitation with a mixture of saturating amounts of monoclonal antibodies against the ha epitope (Babco, Berkeley, CA) and βgal (Promega, Madison, WI). The samples were incubated at 4°C for 2 h, with rotation, followed by the addition of 10 µl of protein A–Sepharose suspension (Repligen, Cambridge, MA), further incubation for 1 h, and centrifugation at 12 000 *g* for 30 s. The immunoprecipitates were washed four times with 0.8 ml of the lysis buffer plus 0.1% SDS, resuspended in SDS-sample buffer (Ausubel *et al.*, 1996), heated at 100°C for 3 min and fractionated by SDS–10% PAGE, followed by autoradiography and quantitation with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA), using the reference provided by the UPR technique (Figures 2 and 5) (Lévy *et al.*, 1996).

Pulse–chase assays with JD54 cells, in which *UBR1 was* expressed from the P_{GAL1} promoter, were carried out by growing 20 ml cultures to A_{600} 0.5–1 in SR(–Trp), collecting and resuspending the cells in 0.5 ml of the same medium containing 0.2 mM $CuSO_4$ and labeling with 0.3 mCi of [^{35}S]methionine/cysteine for 10 min at 30°C. The cells were washed twice with SR(–Trp), and transferred to 0.5 ml of SR (–Trp) containing 10 mM L-methionine and 1 mM cysteine. After a further 20 min incubation at 30°C, galactose was added (to a final concentration of 3%) to induce the expression of *UBR1 and* initiate the chase, whose time points are indicated in Figures 2D and 3D.

ID^5, the extent of initial decay at the end of the 5 min pulse, was calculated as follows: $ID^5 = \{1 - [X\beta gal]_0/[Mete^K\beta gal]_0\} \times 100\%$. This parameter (the upper index refers to the length of pulse) equals 100% minus the ratio of ^{35}S in an X-βgal to ^{35}S in the reference protein dha-Ub (DHFR-ha-Ub^{R48}) at the end of the pulse, normalized against the same ratio with metabolically stable Met-e^K-βgal (Lévy *et al.*, 1996). To denote the observed half-lives of a test protein at different regions of a non-exponential decay curve, a generalized half-life term $t_{0.5}^{y-z}$ was used, in which 0.5 denotes the parameter's half-life aspect and *y–z* denotes the relevant time interval, from *y* to *z* min of chase.

Acknowledgements

We thank the current and former members of the Varshavsky laboratory, particularly F.Lévy, A.Webster and Y.Xie, for helpful discussions. We also thank F.Du, J.Sheng, H.-R.Wang, Y.Xie, H.Rao and especially G.Turner for comments on the manuscript. This work was supported by grants to A.V. from the National Institutes of Health (DK39520) and the US Army Breast Cancer Research Program (DAMD179818042).

References

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Smith,J.A., Seidman,J.G. and Struhl,K. (eds) (1996) *Current Protocols in Molecular Biology.* Wiley-Interscience, New York, NY.

Bachmair,A. and Varshavsky,A. (1989) The degradation signal in a short-lived protein. *Cell*, **56**, 1019–1032.

Bachmair,A., Finley,D. and Varshavsky,A. (1986) *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science*, **234**, 179–186.

Baker,R.T. and Varshavsky,A. (1991) Inhibition of the N-end rule pathway in living cells. *Proc. Natl Acad. Sci. USA*, **87**, 2374–2378.

Baker,R.T. and Varshavsky,A. (1995) Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway. *J. Biol. Chem.*, **270**, 12065–12074.

Baumeister,W., Walz,J., Zühl,F. and Seemüller,E. (1998) The proteasome: paradigm of a self-compartmentalizing protease. *Cell*, **92**, 367–380.

Chau,V., Tobias,J.W., Bachmair,A., Marriott,D., Ecker,D.J., Gonda,D.K. and Varshavsky,A. (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science*, **243**, 1576–1583.

Clarke,N.D. (1995) Sequence 'minimization': exploring the sequence landscape with simplified sequences. *Curr. Opin. Biotechnol.*, **6**, 467–472.

Coux,O., Tanaka,K. and Goldberg,A.L. (1996) Structure and functions of the 20S and 26S proteasomes. *Annu. Rev. Biochem.*, **65**, 801–817.

deGroot,R.J., Rümenapf,T., Kuhn,R.J. and Strauss,J.H. (1991) Sindbis virus RNA polymerase is degraded by the N-end rule pathway. *Proc. Natl Acad. Sci. USA*, **88**, 8967–8971.

Dohmen,R.J., Wu,P. and Varshavsky,A. (1994) Heat-inducible degron: a method for constructing temperature-sensitive mutants. *Science*, **263**, 1273–1276.

Ghislain,M., Dohmen,R.J., Lévy,F. and Varshavsky,A. (1996) Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in *Saccharomyces cerevisiae. EMBO J.*, **15**, 4884–4899.

Grigoryev,S., Stewart,A.E., Kwon,Y.T., Arfin,S.M., Bradshaw,R.A., Jenkins,N.A., Copeland,N.G. and Varshavsky,A. (1996) A mouse amidase specific for N-terminal asparagine. The gene, the enzyme and their function in the N-end rule pathway. *J. Biol. Chem.*, **271**, 28521–28532.

Hershko,A. and Ciechanover,A. (1998) The ubiquitin system. *Annu. Rev. Biochem.*, **76**, 425–479.

Hill,C.P., Johnston,N.L. and Cohen,R.E. (1993) Crystal structure of a ubiquitin dependent degradation substrate: a three-disulfide form of lysozyme. *Proc. Natl Acad. Sci. USA*, **90**, 4136–4140.

Hilt,W. and Wolf,D.H. (1996) Proteasomes: destruction as a programme. *Trends Biochem. Sci.*, **21**, 96–102.

Hochstrasser,M. (1996) Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.*, **30**, 405–439.

Johnson,E.S., Gonda,D.K. and Varshavsky,A. (1990) *Cis-trans* recognition and subunit-specific degradation of short-lived proteins. *Nature*, **346**, 287–291.

Johnson,E.S., Bartel,B.W. and Varshavsky,A. (1992) Ubiquitin as a degradation signal. *EMBO J.*, **11**, 497–505.

Koepp,D.M., Harper,J.W. and Elledge,S.J. (1999) How the cyclin became a cyclin: regulated proteolysis in the cell cycle. *Cell*, **97**, 431–434.

Kwon,Y.T. *et al.* (1998) The mouse and human genes encoding the recognition component of the N-end rule pathway. *Proc. Natl Acad. Sci. USA*, **95**, 7898–7903.

Kwon,Y.T., Kashina,A.S. and Varshavsky,A. (1999) Alternative splicing results in differential expression, activity and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway. *Mol. Cell. Biol.*, **19**, 182–193.

Laney,J.D. and Hochstrasser,M. (1999) Substrate targeting in the ubiquitin system. *Cell*, **97**, 427–430.

Lévy,F., Johnsson,N., Rumenapf,T. and Varshavsky,A. (1996) Using ubiquitin to follow the metabolic fate of a protein. *Proc. Natl Acad. Sci. USA*, **93**, 4907–4912.

Lévy,F., Johnston,J.A. and Varshavsky,A. (1999) Analysis of a conditional degradation signal in yeast and mammalian cells. *Eur. J. Biochem.*, **259**, 244–252.

Madura,K. and Varshavsky,A. (1994) Degradation of Ga by the N-end rule pathway. *Science*, **265**, 1454–1458.

Madura,K., Dohmen,R.J. and Varshavsky,A. (1993) N-recognin/Ubc2 interactions in the N-end rule pathway. *J. Biol. Chem.*, **268**, 12046–12054.

Oliphant,A.R., Nussbaum,A.L. and Struhl,K. (1986) Cloning of random-sequence oligodeoxynucleotides. *Gene*, **44**, 177–183.

Peters,J.M. (1998) SCF and APC: the Yin and Yang of cell cycle regulated proteolysis. *Curr. Opin. Cell Biol.*, **10**, 759–768.

Pickart,C.M. (1997) Targeting of substrates to the 26S proteasome. *FASEB J.*, **11**, 1055–1066.

Rechsteiner,M. (1998) The 26S proteasome. In Peters,J.M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell.* Plenum Press, New York, NY, pp. 147–189.

Sadis,S.C.A. and Finley,D. (1995) Synthetic signals for ubiquitin-dependent proteolysis. *Mol. Cell. Biol.*, **15**, 4086–4094.

Scheffner,M., Smith,S. and Jentsch,S. (1998) The ubiquitin conjugation system. In Peters,J.-M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell.* Plenum Press, New York, NY, pp. 65–98.

Sijts,A.J., Pilip,I. and Pamer,E.G. (1997) The *Listeria monocytogenes*-secreted p60 protein is an N-end rule substrate in the cytosol of infected cells. Implications for major histocompatibility complex class I antigen processing of bacterial proteins. *J. Biol. Chem.*, **272**, 19261–19268.

Stewart,A.E., Arfin,S.M. and Bradshaw,R.A. (1995) The sequence of porcine protein NH_2-terminal asparagine amidohydrolase. A new component of the N-end rule pathway. *J. Biol. Chem.*, **270**, 25–28.

Tobery,T. and Siliciano,R.F. (1999) Induction of enhanced CTL-dependent protective immunity *in vivo* by N-end rule targeting of a model tumor antigen. *J. Immunol.*, **162**, 639–642.

Townsend,A., Bastin,J., Gould,K., Brownlee,G., Andrew,M., Coupar,B., Boyle,D., Chan,S. and Smith,G. (1988) Defective presentation to class I-restricted cytotoxic T lymphocytes in vaccinia-infected cells is overcome by enhanced degradation of antigen. *J. Exp. Med.*, **168**, 1211–1224.

Tyers,M. and Willems,A.R. (1999) One ring to rule a superfamily of E3 ubiquitin ligases. *Science*, **284**, 602–604.

Varshavsky,A. (1996) The N-end rule: functions, mysteries, uses. *Proc. Natl Acad. Sci. USA*, **93**, 12142–12149.

Varshavsky,A. (1997) The ubiquitin system. *Trends Biochem. Sci.*, **22**, 383–387.

Wilkinson,K. and Hochstrasser,M. (1998) The deubiquitinating enzymes. In Peters,J.-M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell.* Plenum Press, New York, NY.

Worley,C.K., Ling,R. and Callis,J. (1998) Engineering *in vivo* instability of firefly luciferase and *Escherichia coli* beta-glucuronidase in higher plants using recognition elements from the ubiquitin pathway. *Plant Mol. Biol.*, **37**, 337–347.

Received August 18, 1999; revised September 15, 1999; accepted September 16, 1999

Curr Genet (1999) 36: 113–123

© Springer-Verlag 1999

ORIGINAL PAPER

Youming Xie · Alexander Varshavsky

The N-end rule pathway is required for import of histidine in yeast lacking the kinesin-like protein Cin8p

Received: 7 April / 17 May 1999

Abstract The N-end rule pathway is a ubiquitin-dependent proteolytic system whose targets include proteins bearing destabilizing N-terminal residues. We carried out a synthetic lethal screen for *Saccharomyces cerevisiae* mutants that require the N-end rule pathway for cell viability. A mutant thus identified, termed *sln2*, could not grow in the absence of Ubr1p, the recognition component of the N-end rule pathway, which was not essential for viability of the parental strain under the same conditions. Further analysis showed that inviability of *sln2 ubr1Δ* cells could be rescued either by the *HIS3* gene (which was absent from the parental strain) or by a high concentration of histidine in the medium. This defect in histidine uptake, exhibited by the *sln2* mutant in the absence but not in the presence of Ubr1p, was traced to the gene *HIP1*, which encodes the histidine transporter. *HIP1* was underexpressed in *sln2 ubr1Δ* cells, in comparison to either *sln2 UBR1* or *SLN2 ubr1Δ* cells. Yet another property of the *sln2* mutant was its inviability at 37 °C, which could not be rescued by either *UBR1* or *HIS3*. This feature of *sln2* allowed the cloning of *SLN2*, which was found to be a gene called *CIN8*, encoding a kinesin-like protein. Thus, either the N-end rule pathway or Cin8p must be present for the viability-sustaining rate of histidine import in *S. cerevisiae* auxotrophic for histidine. We consider possible mechanisms of this previously unsuspected link between kinesins, ubiquitin-dependent proteolysis, and the import of histidine.

Key words Ubiquitin · Proteolysis · Kinesin · Histidine import · *CIN8* · *HIP1* · *UBR1*

Introduction

A number of regulatory circuits involve metabolically unstable proteins. A short in vivo half-life of a regulator provides a way to generate its spatial gradients and allows for rapid adjustments of its concentration, or subunit composition, through changes in the rate of regulator's synthesis or degradation. Short in vivo half-lives are also characteristic of damaged or otherwise abnormal proteins (Hochstrasser 1996; Varshavsky 1997; Peters et al. 1998; Scheffner et al. 1998). Features of proteins that confer metabolic instability are called degradation signals, or degrons. The essential component of one degradation signal, the first to be identified, is a destabilizing N-terminal residue of a protein (Bachmair et al. 1986; Varshavsky 1996). The set of amino-acid residues which are destabilizing in a given cell yields a rule, called the N-end rule, which relates the in vivo half-life of a protein to the identity of its N-terminal residue. Similar but distinct N-end rule pathways are present in all organisms examined, from mammals and plants to fungi and bacteria (Bachmair and Varshavsky 1989; Gonda et al. 1989; Tobias et al. 1991; Varshavsky 1996; Worley et al. 1998).

In eukaryotes, the degradation signals (N-degrons) recognized by the N-end rule pathway comprise two determinants: a destabilizing N-terminal residue and an internal lysine or lysines (Bachmair and Varshavsky 1989; Hill et al. 1993; Dohmen et al. 1994). The Lys residue is the site of formation of a multiubiquitin chain (Chau et al. 1989). The N-end rule pathway is thus one pathway of the ubiquitin system. Ubiquitin is a 76-residue protein whose covalent conjugation to other proteins plays a role in a multitude of processes, including cell growth, division, differentiation, and responses to stress (Johnson et al. 1995; Hochstrasser 1996; Hershko 1997; Pickart 1997; Varshavsky 1997; Scheffner et al. 1998). In most of these processes, Ub acts through routes that involve the degradation of ubiquitin-protein conjugates by the 26S proteasome, an ATP-dependent

---

Y. Xie · A. Varshavsky (✉)
Division of Biology, 147-75, Caltech, 1200 East California Blvd., Pasadena, CA 91125, USA
e-mail: avarsh@cco.caltech.edu
Tel.: +1-626-395 3785; Fax: +1-626-440 9821

Communicated by C.P. Hollenberg

multisubunit protease (Rechsteiner et al. 1993; Coux et al. 1996; Baumeister et al. 1998; Glickman et al. 1998).

The N-end rule is organized hierarchically. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are *tertiary* destabilizing N-terminal residues in that they function through their conversion into the *secondary* destabilizing N-terminal residues Asp and Glu. This conversion is mediated by the *NTA1*-encoded N-terminal amidase (Nt-amidase) (Baker and Varshavsky 1995). The destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *ATE1*-encoded Arg-tRNA-protein transferase (R-transferase), to Arg, one of the *primary* destabilizing residues (Balzi et al. 1990). The primary destabilizing N-terminal residues are bound directly by the *UBR1*-encoded N-recognin (also called E3), the recognition component of the N-end rule pathway (Bartel et al. 1990). In *S. cerevisiae*, N-recognin is a 225 kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues – Phe, Leu, Trp, Tyr, Ile, Arg, Lys, and His (Varshavsky 1996). The *Ubr1*-encoded mammalian N-recognin (E3α) has also been characterized (Hershko 1991; Kwon et al. 1998). Ubr1p has at least two substrate-binding sites. The type-1 site is specific for the basic N-terminal residues Arg, Lys and His. The type-2 site is specific for the bulky hydrophobic N-terminal residues Phe, Leu, Trp, Tyr, and Ile (Varshavsky 1996). Both yeast and mammalian Ubr1p have also been shown to recognize a set of internal degrons that remain to be characterized (Gonen et al. 1991; Madura and Varshavsky 1994; Byrd et al. 1998).

The known functions of the N-end rule pathway include the control of peptide import in *S. cerevisiae*, through the degradation of Cup9p, a transcriptional repressor of *PTR2* which encodes the peptide transporter (Alagramam et al. 1995; Byrd et al. 1998); a role in regulating the Sln1p-dependent phosphorylation cascade that mediates osmoregulation in *S. cerevisiae* (Ota and Varshavsky 1993); the degradation of Gpa1p, a Gα protein of *S. cerevisiae* (Madura 1994, #60); and the conditional degradation of alphaviral RNA polymerase in virus-infected metazoan cells (deGroot et al. 1991; Varshavsky et al. 1998). Physiological N-end rule substrates were also identified among the proteins secreted into the mammalian cell's cytosol by intracellular parasites such as the bacterium *Listeria monocytogenes* (Sijts et al. 1997). Inhibition of the N-end rule pathway was reported to interfere with mammalian cell differentiation (Hondermarck et al. 1992), and to delay limb regeneration in amphibians (Taban et al. 1996). In mammals, the N-end rule pathway is also likely to play a role in catabolic states that result in muscle atrophy (Solomon et al. 1998 a,b).

While the absence of the N-end rule pathway (specifically, the absence of Ubr1p) from *S. cerevisiae* is lethal under conditions where the import of peptides is essential for cell survival (Alagramam et al. 1995; Byrd et al. 1998), *ubr1Δ* cells are viable and only slightly growth-retarded on standard media (Bartel et al. 1990). Synthetic lethal screens can often identify pathways whose essential functions overlap (or interact) with the function of a system of interest (Singer-Kruger and Ferro-Novick 1997; Ho et al. 1998; Merrill and Holm 1998). An earlier screen for *S. cerevisiae* mutants that require Ubr1p for viability has identified Sln1p, which contains a sensor kinase and response regulator domains analogous to those of the prokaryotic two-component regulators (Ota and Varshavsky 1993). Sln1p functions in the HOG signaling pathway, which regulates the response of cells to changes in the media osmolarity (Posas et al. 1996). *sln1* mutants are growth-impaired in presence of Ubr1p and inviable in its absence (Ota and Varshavsky 1993). The identity of a relevant Ubr1p substrate in the HOG pathway is unknown.

In the present work, we carried out a modified synthetic lethal screen for *S. cerevisiae* mutants that require Ubr1p for viability. Our findings uncovered a previously unsuspected connection between kinesins, the N-end rule pathway, and the import of histidine.

Materials and methods

Yeast strains, media, and genetic procedures. *S. cerevisiae* strains used in this study are listed in Table 1. Yeast were grown in rich (YPD) medium, or in synthetic media containing 0.67% yeast nitrogen base without amino acids (Difco), auxotrophic nutrients, including adenine, arginine, histidine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan and uracil, and either 2% dextrose (glucose) (SD medium) or 2% galactose (SG medium) (Sherman et al. 1986). In some experiments, the concentration of histidine in SD medium was increased 10-fold, to 0.2 mg/ml. Unless stated otherwise, cells were incubated at 30 °C. Yeast mating, sporulation, and tetrad dissection were performed as described (Sherman et al. 1986). The mating-type switch was carried out through transformation with the plasmid pGAL-HO (Herskowitz and Jensen 1991). *S. cerevisiae* were transformed as described by Chen et al. (1992). The *Escherichia coli* strain used was DH5α (Ausubel et al. 1996).

Synthetic lethal screen. The haploid P_{GAL1}-*UBR1* strain JD54, in which the *UBR1* gene is expressed from the P_{GAL1} promoter (Ghislain et al. 1996), was mutagenized with ethyl methanesulfonate (EMS) (Sigma, St. Louis, Mo.) to approximately 7% survival upon plating to SG medium at 30 °C. The colonies on SG plates were replica-plated onto SD plates. The colonies that grew on SG but not on SD media were re-tested for the absence of growth on SD plates. The candidate isolates were transformed with plasmids that expressed *UBR1* either from its natural promoter [pUBR1 (Bartel et al. 1990)] or the much stronger P_{ADH1} promoter, and tested for survival on SD plates. One mutant, termed *sln2* (AVY200), was rescued by *UBR1* on SD plates, and was selected for further study.

Isolation of HIS3 as a suppressor of sln2. The *sln2* mutant was transformed with a yeast genomic DNA library (American Type Culture Collection, #77164) in the *TRP1*, *CEN6*-based vector pRS200 (Sikorski and Hieter 1989). Cells were plated on SD(-Trp) plates and incubated at 30 °C for 2–3 days. Out of approximately 4×10^4 transformants, 25 could grow on SD(-Trp). Library plasmids were rescued from these transformants and re-transformed into *sln2* cells. Eight of these plasmids conferred viability on *sln2* cells after re-transformation and selection on SD(-Trp). Restriction analyses showed that these plasmids were either identical or had

Table 1 *S. cerevisiae* strains used in this study

Strain	Genotype	References
JD52	*MAT***a** *ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112*	Johnson et al. (1995)
JD54	*MAT***a** *ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112 GAL1::UBR1*	Ghislain et al. (1996)
AVY199	*MATα ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112 GAL1::UBR1*	Derivative of JD54
AVY200[a]	*MAT***a** *sln2 ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112 GAL1::UBR1*	Derivative of JD54
AVY201	*MAT***a** *ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112 GAL1::UBR1 hip1Δ-3::URA3*	Derivative of JD54
AVY202	*MAT***a** *ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112 GAL1::UBR1 cin8Δ-6::URA3*	Derivative of JD54
AVY203	*MAT***a** *ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112 cin8Δ-10::URA3*	Derivative of JD52
AVY204	*MAT***a** *sln2 ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3, 112 GAL1::UBR1 cup9Δ-1::LEU2*	Derivative of AVY200

[a] Mutant (*sln2*) that requires UBR1 for viability

large fragments in common. Truncation and sequencing analyses localized the complementing activity to a 1.7-kb *Bam*HI fragment containing the *HIS3* gene. This finding was independently confirmed by showing that the *HIS3* open reading frame (ORF), expressed from the P_{MET25} promoter (Mumberg et al. 1994), was also able to complement the *sln2* mutant on SD plates.

Histidine uptake assay. This assay was carried out as described (Grauslund et al. 1995), with minor modifications. Briefly, the strains to be assayed were grown to A_{600} of about 0.6 in SD(-His) medium at 30°C. Cells were harvested from 5-ml cultures and resuspended in 0.4–0.5 ml of SD(-His) at an identical cell density, adjusted using a hemocytometer. For each sample, 0.1 ml of suspension was added to 0.1 ml of SD(-His) medium containing 8 μM of L-[^{14}C]histidine (Amersham, Arlington Heights, Ill.; specific radioactivity 0.3 Ci/mmol), and incubated at 30°C for 2 min. Samples were removed, filtered through a #30 glass-fiber filter (Schleicher and Schuell), washed with 10 ml of ice-cold water, and air-dried. ^{14}C radioactivity was determined using a liquid scintillation counter. Since high concentrations of histidine in the medium would have interfered with the assay, *HIS3* was introduced (as the pRS313 plasmid) into all *his3Δ* strains to be tested, making them prototrophic for histidine.

Northern hybridization. S. cerevisiae strains were grown at 30°C to A_{600} of about 1. RNA was isolated as described (Schmitt et al. 1990) and fractionated by electrophoresis in formaldehyde-containing agarose gels (Ausubel et al. 1996). Briefly, approximately 40 μg of RNA per lane was fractionated in a 1.2% agarose gel containing 0.66 M formaldehyde, 1 × MOPS buffer and 0.1 μg/ml of ethidium bromide (Ausubel et al. 1996). Electrophoresis was performed at 5 V/cm in 1 × MOPS buffer. Fractionated RNA was transferred onto a GeneScreen Plus membrane (DuPont NEN, Boston, Mass.) according to the manufacturer's instructions. DNA probes were labeled with [α-^{32}P]dCTP using a DNA labeling kit (Pharmacia Biotech, Piscataway, N.J.). Hybridization was carried out using a NorthernMax kit (Ambion, Austin, Tex.). Hybridization patterns were detected by autoradiography and quantitated using PhosphorImager (Molecular Dynamics).

hip1Δ mutant. This mutant was produced from JD54 (Table 1) using a direct deletion technique described previously (Baudin et al. 1993). Briefly, two oligodeoxynucleotides of 66 nt were synthesized, attagtacacaaatacagatatacaacctcaacatcaaaaatgtcgaaagctacatataaggaacg and gatatggaaattccgctgtatagctcatctctttcccttcattagttttgctggcctcatcttctc. One oligonucleotide comprised 40 nt of a sequence derived from the 5′ untranslated region (UTR) of the *HIP1* gene, followed by 26 nt of a sequence derived from the *URA3* ORF, beginning from its start codon. The second oligonucleotide comprised 40 nt of a sequence derived from the 3′ UTR of *HIP1*, followed by 26 nt of a sequence derived from the 3′-proximal region of the *URA3* ORF, including its stop codon. The two oligonucleotides were used as PCR primers to amplify *URA3* from the pRS306 vector (Sikorski and Hieter 1989). The resulting targeting fragment contained *URA3* flanked on either side by 40-bp sequences derived from, respectively, the 5′ and 3′ UTRs of *HIP1*. JD54 cells (Table 1) were transformed with the targeting fragment, and selected for Ura$^+$ transformants at 30°C on SD(-Ura) plates containing 0.2 mg/ml of histidine. In a typical experiment, approximately 50% of Ura$^+$ transformants were the expected *hip1Δ* mutants, as verified by PCR, and also by checking for the phenotype of dependence on high concentrations of histidine in the medium.

Isolation of HIP1. A gap-repair method was used to rescue *HIP1* from both wild-type (JD54) and *sln2* (AVY200) strains (Rothstein 1991). Two fragments (the left and right arms) flanking the gap were amplified by PCR from either JD54 or *sln2* genomic DNA. The left arm was a 693-bp *Xho*I-*Hin*dIII fragment corresponding to the 5′ UTR of *HIP1* from −976 bp to −284 bp. The right arm was a 1044-bp *Hin*dIII-*Xba*I fragment, corresponding to the 3′ UTR of *HIP1*, 470 bp downstream from the stop codon of *HIP1*. The left and right arms were inserted into the *Xho*I and *Xba*I sites of pRS315, a *LEU2*, *CEN6*-based vector (Sikorski and Hieter 1989). The resulting construct was linearized with *Hin*dIII, and transformed into wild-type and *sln2* cells (see Fig. 3 A). Plasmids were isolated from the Leu$^+$ transformants. Gap-filled plasmids (pRS315HIP1) were identified by restriction analysis, and also by sequencing the relevant junctional regions. The primers for producing the left-arm fragment were ggggctgaaacacaagaattcataatg and gcacctacttcaactccttagcag. The initial PCR product was digested with *Xho*I and *Hin*dIII, yielding the left-arm DNA fragment. The primers for producing the right arm were ctactacagatattacatgtggcg and ggtttcggtagaatcggtagattg. The initial PCR product was digested with *Hin*dIII and *Xba*I, yielding the right arm. All PCR products were verified by DNA sequencing.

Construction of the sln2 cup9Δ mutant. The previously described *cup9::LEU2* disruption allele (Byrd et al. 1998) was used to replace the *CUP9* allele in *sln2* cells through homologous recombination (Rothstein 1991), yielding the strain AVY204 (Table 1). The absence of Cup9p was verified by testing for the ability of AVY204 cells to import dipeptides in the absence of *UBR1* (on SD plates) (Byrd et al. 1998), and also by PCR analysis.

Isolation of the SLN2 (CIN8) gene. The *S. cerevisiae* genomic DNA library used for the isolation of *HIS3* (see above) was transformed into the *sln2* strain (AVY200) (Table 1). Cells were plated on SD(-Trp) plates containing histidine at 0.2 mg/ml, and incubated at 30°C for 8 h. The temperature was then raised to 37°C, until the colonies formed (about 3 days). Out of approximately 6×10^4 transformants at 30°C, 15 could grow on SD(-Trp, +0.2 mg/ml His) at 37°C. Library derived plasmids were rescued from these 15 transformants and re-transformed into *sln2* cells, with selection on SD(-Trp) plates at 37°C. Three plasmids conferred viability on *sln2* cells after re-transformation. Restriction mapping showed the inserts of these plasmids to be identical. Truncation and sequence analyses confirmed that the complementing activity was confined to the *CIN8* locus.

Isolation of the CIN8 allele from sln2 cells.* A gap-repair method (Rothstein 1991) described above for *HIP1* was used to isolate the mutant allele of *CIN8* (denoted as *CIN8**) from *sln2* cells. A 4973-bp *Eco*RI-*Bam*HI fragment containing wild-type *CIN8*, including its 5′ (464 bp) and 3′ (1353 bp) untranslated regions (UTRs), was subcloned into the *TRP1*-based vector pRS314 (Sikorski and Hieter 1989). The resulting pRS314CIN8 was gapped by digestion with *Pac*I and *Bse*RI, and transformed into the *sln2* strain (AVY200). Plasmids were isolated from the Trp$^+$ transformants. The gap-filled plasmids (pRS314CIN8*) were verified by restriction analysis and sequencing of the relevant junctional regions.

Construction of null cin8 mutants. An approximately 2.5-kb *Sph*I-*Aat*II DNA fragment of pRS314CIN8 was replaced by a PCR-produced, approximately 1.1-kb, URA3-containing *Sph*I-*Aat*II fragment of YIp5, yielding pRS314cin8::URA3. A *Sal*I-*Bam*HI fragment of pRS314cin8::URA3 containing the *cin8*::URA3 deletion/disruption allele was used to replace, through homologous recombination, the *CIN8* gene in JD52, a wild-type (Ubr1$^+$) strain, and in JD54, a P_{GAL1}::*UBR1* strain. The resulting *cin8Δ* mutants AVY203 and AVY202, respectively (Table 1), lacked codons 169–1015 of the *CIN8* ORF. The structure of the *cin8Δ* allele, and the temperature sensitivity of *cin8Δ* mutants were confirmed, respectively, using PCR and growth tests at 37°C. The primers for producing the *URA3*-containing DNA fragment were cattGCATGcgtccgtggaattctcatgtttgac and agtgGACGTCtaccaggtgcctgactgcgttagc (uppercase letters denote the *Sph*I and *Aat*II sites, respectively).

Results

Isolation of the *sln2* mutant

The earlier screen for *S. cerevisiae* mutants requiring the N-end rule pathway (specifically, Ubr1p) for viability (Ota and Varshavsky 1993) was far from exhaustive, and in addition utilized wild-type levels of Ubr1p expression. Since the activity of the N-end rule pathway is higher in cells overproducing Ubr1p (Bartel et al. 1990; Madura and Varshavsky 1994), the present screen was carried out using the strain JD54 (Ghislain et al. 1996), in which *UBR1* was expressed from the galactose-inducible, dextrose-repressible P_{GAL1} promoter (Table 1). Mutagenized JD54 cells were plated onto SG (galactose-containing) plates, and the colonies formed were replica-plated onto SD (dextrose-containing) plates, where the N-end rule pathway was inactive, owing to the absence of Ubr1p expression. Cells that yielded colonies on SG but not on SD plates were putative *sln* (*s*ynthetic *l*ethal of *N*-end rule) mutants. Four such isolates were identified among approximately 3.5×10^5 colonies screened.

To verify that the conditional viability of these isolates was in fact dependent on the presence of Ubr1p, rather than on the nature of carbon source (dextrose versus galactose), the putative mutants were transformed with a low-copy plasmid pUBR1 expressing *UBR1* from its natural promoter, and plated on SD plates. Only one isolate, termed *sln2* (strain AVY200; Table 1), formed colonies on SD plates in the presence of pUBR1 but not in the presence of the control vector (Fig. 1 A–C). The inviability of *sln2* cells on SD plates

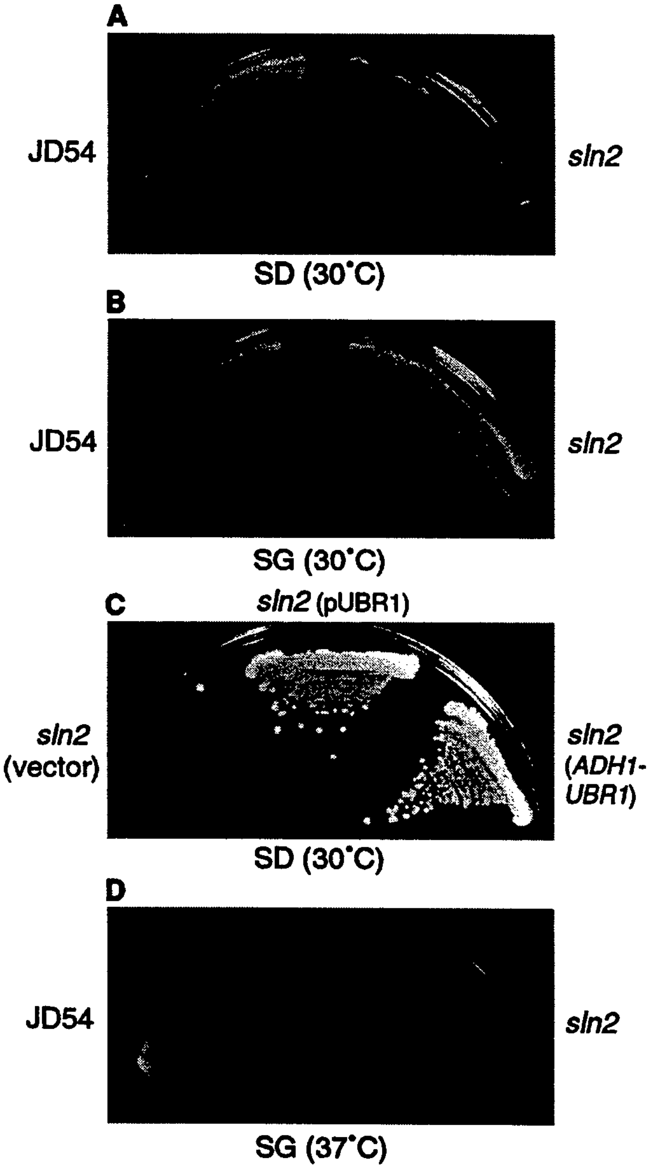


Fig. 1 A–D A *S. cerevisiae* mutant whose viability requires the N-end rule pathway. (**A** and **B**) the AVY200 (*sln2* P_{GAL}-*UBR1*) strain grows on SG but not on SD plates (see Results). (**C**) evidence that the viability of *sln2* mutant requires Ubr1p rather than galactose. AVY200 (*sln2* P_{GAL}-*UBR1*) cells were streaked on SD plates after having been transformed either with pUBR1, a low-copy plasmid expressing *UBR1* from its natural promoter or with pADH-UBR1, another *CEN*-based plasmid expressing *UBR1* from the constitutive and strong P_{ADH1} promoter, or with vector alone. (**D**) *sln2* cells are *ts* for growth. The parental JD54 (P_{GAL}-*UBR1*) strain and the AVY200 (*sln2* P_{GAL}-*UBR1*) strain (Table 1) were streaked on SG plates and incubated at 37 °C

could be rescued by *UBR1* expressed at either normal (pUBR1) or much higher levels (pRB208, which expressed *UBR1* from the P_{ADH1} promoter) (Fig. 1 C).

Histidine import defect of *sln2* cells in the absence of Ubr1p

Crossing AVY200 (*sln2*) to the strain AVY199 (congenic to JD54 but of the opposite mating type) (Table 1) indicated that the *sln2* mutation was recessive (data not shown). We used a library of *S. cerevisiae* genomic DNA fragments in a low-copy vector to transform AVY200 cells, screening for genes that could rescue the inviability of *sln2* on SD plates (in the absence of Ubr1p). Surprisingly, *HIS3*, a gene of the histidine biosynthesis pathway, was found to be a strong suppressor of the lethality of AVY200 (*sln2*) on SD plates (Fig. 2 A). Since AVY200 is a *his3Δ* strain (Table 1), *HIS3* is, by definition, an extragenic suppressor of the conditionally lethal *sln2* phenotype, i.e., *HIS3* is distinct from *SLN2*.

The original JD54 (*his3Δ*) strain, being a histidine auxotroph, required the presence of histidine in the medium. In the presence of *HIS3*, the JD54-derived AVY200 (*sln2*) mutant would be expected to become prototrophic for histidine. Could the effect of the *sln2* mutation be an impairment of the histidine uptake in the *absence* of the N-end rule pathway (specifically, in the absence of Ubr1p)? This interpretation could account for the suppression of inviability of AVY200 cells on SD plates by the *HIS3* gene, inasmuch as the resulting strain would be a histidine prototroph. We first asked whether increasing the concentration of histidine in the growth medium would rescue AVY200 (*sln2*) cells on SD plates. [It has previously been shown that cells defective in histidine uptake could grow on a medium containing sufficiently high concentrations of histidine (Tanaka and Fink 1985).] Indeed, AVY200 (*sln2* P_{GAL}-*UBR1*) cells were found to grow on SD plates containing histidine at 200 µg/ml, but could not grow at the standard histidine concentration of 20 µg/ml (Fig. 2 B).

To examine the histidine import of AVY200 (*sln2*) cells in a different way, we compared the uptake of L-[^{14}C]histidine into *sln2* and other yeast strains. Since high concentrations of histidine in the medium would have interfered with the assay, we introduced the *HIS3* gene into all four of the *his3Δ* strains to be tested, making them prototrophic for histidine. As shown in Fig. 3, in the absence of Ubr1p the relative rate of histidine uptake in AVY200 (*sln2* P_{GAL}-*UBR1*) cells was about 35% of the rate in the parental JD54 (*SLN2* P_{GAL}-*UBR1*) cells. In the presence of Ubr1p (at its wildtype concentration) the rate of histidine uptake in *sln2* cells increased to about 75% of that in *SLN2* cells (Fig. 3), thus accounting, at least in part, for the observed growth of *sln2* cells on SG plates (Fig. 1 B). The incomplete recovery of histidine uptake in Ubr1p-expressing AVY200 (*sln2*) cells was consistent with their

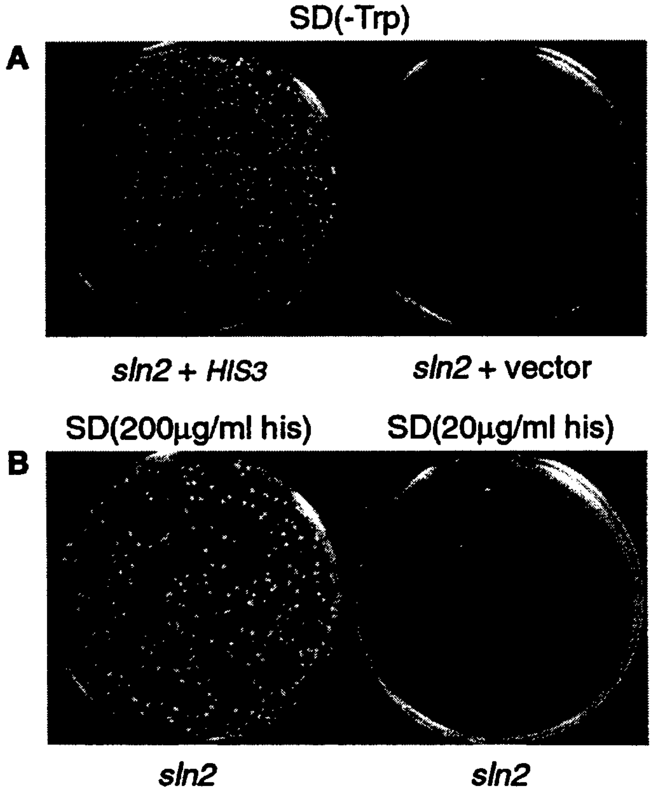


Fig. 2 Inviability of the AVY200 (*sln2* P_{GAL}-*UBR1*) strain on SD plates can be rescued either by the *HIS3* gene (panel A) or by high (>75 µg/ml) concentration of histidine in the medium (panel B). See Results for details

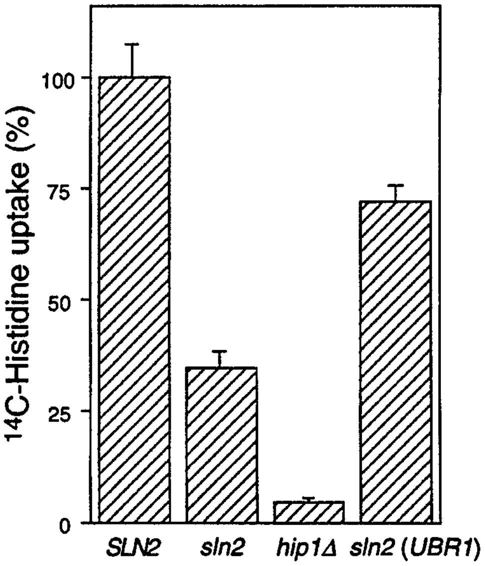


Fig. 3 *sln2* cells are impaired in the import of histidine. *SLN2* is the parental JD54 strain (P_{GAL}-*UBR1*) (Table 1). *hip1Δ* is a *HIP1*-lacking derivative of JD54. *sln2* (pUBR1) is the AVY200 (*sln2* P_{GAL}-*UBR1*) strain carrying pUBR1. The uptake of L-[^{14}C]histidine was measured as described in Materials and methods. The results are expressed as the percent of uptake by *SLN2* cells, and represent the means from three independent experiments. Standard deviations are shown *above* the bars

detectably slower growth on SG plates in comparison to JD54 (*SLN2*) cells (Fig. 1 A,B). Note that the decreased (about 35% of normal) rate of ^{14}C-histidine import in *sln2* cells was still considerably higher than the one in *hip1Δ* cells, which lack the histidine-specific transporter (Fig. 3). This observation is in agreement with the fact that the minimal concentration of histidine that allowed the growth of *sln2* cells on SD plates (75 μg/ml) was much lower than the concentration of histidine (200 μg/ml) required for the comparable growth rate of *hip1Δ* cells (data not shown).

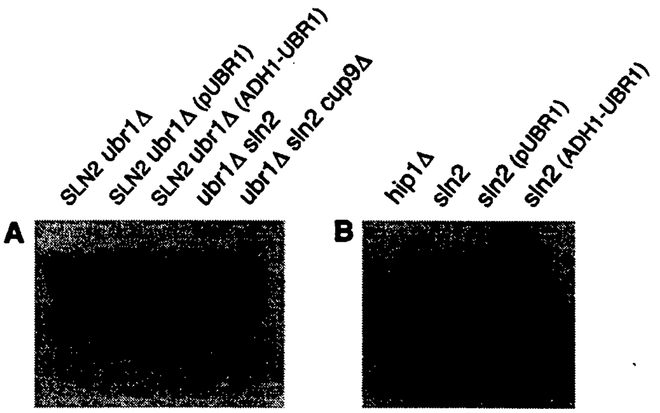


Fig. 5 Northern-hybridization analyses of the *S. cerevisiae HIP1* mRNA in different genetic backgrounds. (**A** and **B**) ^{32}P-labeled *HIP1* and *ACT1* (actin) DNA probes (the latter appropriately diluted) were used to determine the levels of the corresponding mRNAs in electrophoretically fractionated RNA from strains whose relevant genetic backgrounds are indicated at the top of panels. "pUBR1" and "pADH-UBR1" denote RNA from strains carrying low-copy plasmids expressing *UBR1* from either the natural *UBR1* promoter or from the P_{ADH1} promoter. See also Materials and methods

Down-regulation of *HIP1* in the *sln2 ubr1Δ* mutant

In *S. cerevisiae*, the import of histidine across the plasma membrane is carried out largely, if not exclusively, by two transporters (permeases), Hip1p and Gap1p (Tanaka and Fink 1985; Jauniaux and Grenson 1990). Gap1p is a general amino-acid transporter whose activity is repressed by rich nitrogen sources such as NH_4^+ in SD or SG media (Jauniaux and Grenson 1990; Hein et al. 1995). Consequently, the histidine-specific permease Hip1p is the main importer of histidine under these conditions. The reduced uptake of histidine by AVY200 (*sln2*) cells in the absence of Ubr1p (in the absence of the N-end rule pathway) could be caused by a defective *HIP1* allele in *sln2* cells or by a reduced expression of the otherwise wild-type *HIP1*.

To address the first possibility, we used the method of gap repair (Fig. 4 A) (see Materials and methods) to isolate *HIP1* from both AVY200 (*sln2*) and the parental JD54 (*SLN2*) strains. Complementation testing of *HIP1* from these strains with *hip1Δ S. cerevisiae* showed that the two *HIP1* alleles were indistinguishable in their ability to restore histidine import in *hip1Δ* cells (data not shown), strongly suggesting that *sln2* cells contained the wild-type version of *HIP1*. Quantitative Northern hybridization was then used to compare the levels of *HIP1* mRNA in *sln2* and congenic *SLN2* cells in the absence and presence of Ubr1p (Fig. 5 and see Materials and methods). The relative level of *HIP1* mRNA was reproducibly about 2-fold lower in *sln2* cells lacking Ubr1p than in otherwise identical *SLN2* cells or in Ubr1p-containing *SLN2* cells (Fig. 5 A and see Materials and methods). This finding strongly suggested that the inviability of Ubr1p-lacking AVY200 (*sln2* P_{GAL}-*UBR1*) cells at the standard concentration of histidine in SD medium was caused by underexpression of the Hip1p transporter. Consistent with this interpretation, we found that about a 2-fold increase in the *HIP1* copy number, through transformation with a low-copy plasmid expressing *HIP1* from its natural promoter, rescued the inviability of AVY200 cells on SD plates (Fig. 6 A). Note that the *HIS3* gene on a low-copy plasmid rescued AVY200 (*sln2* P_{GAL}-*UBR1*) cells more effectively than did the plasmid-borne extra copy of *HIP1* (Fig. 6 A), again in agreement with the observed down-regulation of *HIP1* in AVY200 cells. This finding also accounted, in hindsight, for the isolation of *HIS3*, but not of *HIP1*, in our (far from exhaustive) complementation screen.

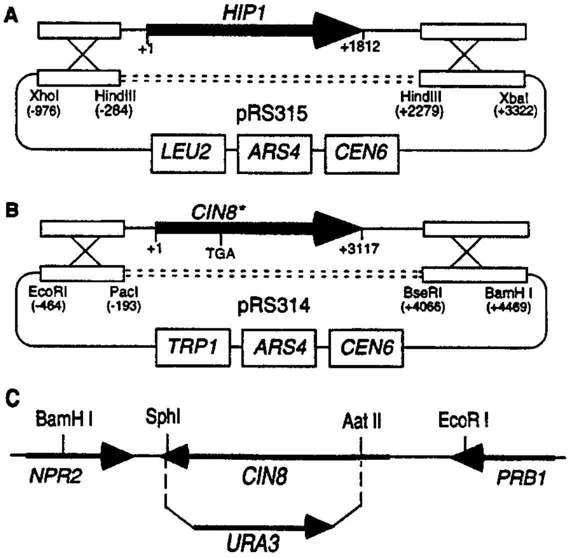


Fig. 4 A–C Gap repair-based cloning of the *HIP1* and *CIN8** alleles from *sln2* cells. Panels **A** and **B** illustrate, respectively, the isolation of *HIP1* and an allele of *CIN8* (*CIN8**) from AVY200 (*sln2* P_{GAL}-*UBR1*) cells. **C** deletion/disruption of the *CIN8* gene. See Materials and methods

Ubr1p increases the expression of *HIP1* in *sln2* but not in *SLN2* cells

The effect of *UBR1* expression on the uptake of histidine by Ubr1p-lacking *sln2* cells (Fig. 3), and the requirement

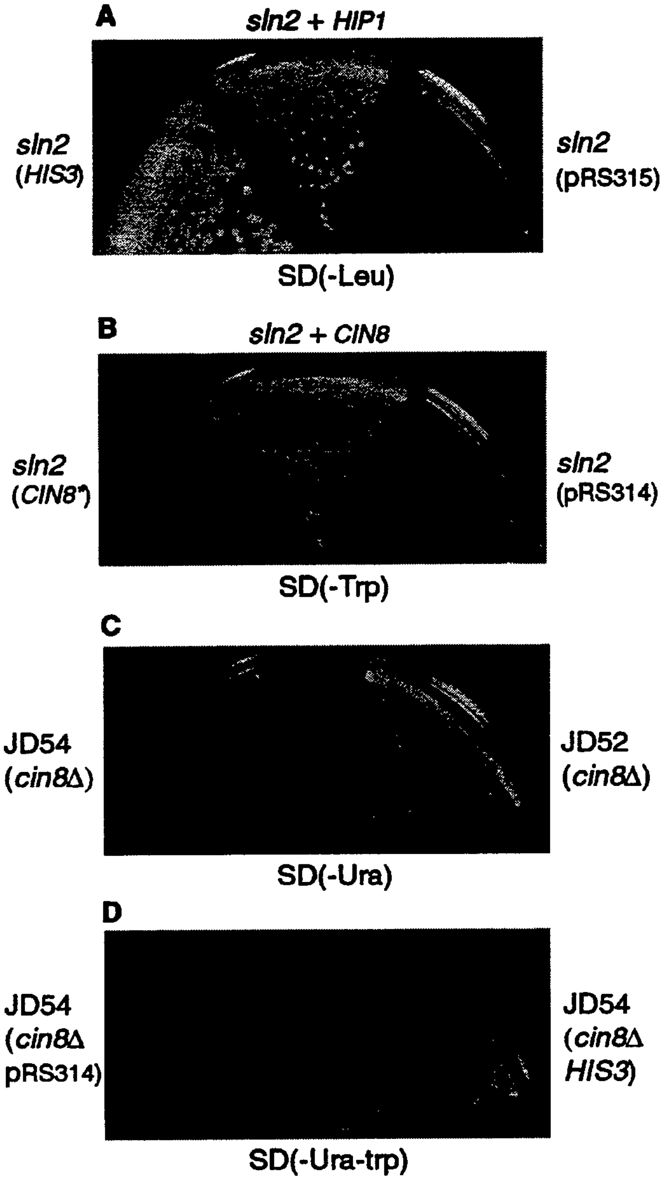


Fig. 6A–D Effects of *HIP1* and *CIN8** on the growth of strains lacking the N-end rule pathway. **A** AVY200 (*sln2* P_{GAL}-*UBR1*) cells (Table 1) carrying a low-copy, *LEU2*-marked vector pRS315 or the otherwise identical plasmids expressing either *HIS3* or *HIP1* were grown on SD(-Leu, +0.2 mg/ml of histidine) plates to allow colony formation by all strains. Individual colonies were then streaked onto SD(-Leu) plates. Note the rescue of the AVY200 (*sln2* P_{GAL}-*UBR1*) cells by *HIP1* and their even stronger rescue by *HIS3*. **B** neither the *CIN8** allele (isolated from *sln2* cells) nor the corresponding *TRP1*-based vector pRS314 rescued inviability of AVY200 (*sln2* P_{GAL}-*UBR1*) cells on SD plates, in contrast to the wild-type *CIN8* gene. **C** *cin8Δ* cells (AVY202 and AVY203, respectively) derived from either the JD54 (P_{GAL1}-*UBR1*) or JD52 (*UBR1*) strains (Table 1) were streaked onto SD plates. Note inviability of *cin8Δ* cells in the absence of *UBR1* expression. Cells were grown at first on SD(-Ura, +0.2 mg/ml of histidine) plates to allow colony formation by both strains. Individual colonies were then streaked onto SD(-Ura) plates. **D** same as in **C** but *cin8Δ* JD54 (*cin8Δ* P_{GAL}-*UBR1*) cells carried either the control vector pRS314 or the same vector expressing *HIS3*. See Materials and methods

of Ubr1p for the viability of these cells on SD plates (Fig. 1 A–C), could be due to an increase of *HIP1* expression in the presence of Ubr1p. Comparisons of *HIP1* mRNA levels among *sln2* strains that either lacked Ubr1p or produced increasing amounts of Ubr1p showed that the expression of *HIP1* was indeed increased in the presence of Ubr1p in *sln2* cells, but not to its wild-type level (Fig. 5). As quantified with PhosphorImager and internal reference, the increase caused by the presence of Ubr1p was, reproducibly, about 1.5-fold, in comparison to the approximately 2-fold increase that would have been required for restoration of the wild-type level of *HIP1* mRNA (Fig. 5 and data not shown). The partial restoration of *HIP1* expression by Ubr1p in *sln2* cells is consistent with the observed partial recovery of histidine uptake (Fig. 3), and with lower than wild-type growth rates of Ubr1p-supplemented *sln2* cells (Fig. 1 B).

We also asked whether the presence of Ubr1p could increase the expression of *HIP1* in *SLN2* cells. As shown in Fig. 5 A, the expression of Ubr1p, at increasing levels, in *SLN2 ubr1Δ* cells did not influence the expression of *HIP1*. Taken together, our results indicated that at least two distinct pathways influence the expression of *HIP1*: a *SLN2*-dependent pathway and the *UBR1*-dependent N-end rule pathway. The latter pathway is sufficient for the viability-maintaining expression of *HIP1* in the absence of *SLN2*. Although the expression of *HIP1* was not extinguished even in the absence of both *SLN2* and *UBR1* (Fig. 5 A), it became low enough to result in growth arrest at the standard concentration of histidine in the medium (Fig. 1 A).

Previous work (Byrd et al. 1998) has shown that the N-end rule pathway strongly up-regulates the expression of *S. cerevisiae* Ptr2p, a peptide transporter, through the degradation of Cup9p, a transcriptional repressor of *PTR2*. We therefore tested whether the Ubr1p-dependent increase of *HIP1* expression in AVY200 (*sln2* P_{GAL}-*UBR1*) cells (Fig. 5) involves the degradation of Cup9p, which is known to either repress or activate several yeast genes (G. Turner, H. Rao, and A.V., unpublished data). A *sln2 cup9Δ* P_{GAL}-*UBR1* strain was constructed; the levels of *HIP1* mRNA were compared between *sln2 CUP9* and *sln2 cup9Δ* strains, and were found to be indistinguishable (Fig. 5 A). Thus, the effect of the N-end rule pathway on the expression of *HIP1* is unlikely to involve Cup9p. These Northern-hybridization data (Fig. 5 A) were also consistent with the observation that a deletion of *CUP9* did not rescue the inviability of AVY200 (*sln2* P_{GAL}-*UBR1*)cells on SD plates (data not shown).

SLN2 is *CIN8*

As described above, both *UBR1* and *HIS3* acted as extragenic suppressors of the *sln2* mutation. To identify the *SLN2* gene, we searched for other phenotypes of

AVY200 (*sln2* P_{GAL}-*UBR1*) cells, and found that they were inviable at 37 °C. The result of crossing AVY200 with the *SLN2* AVY199 strain (Table 1) indicated that the phenotype of temperature sensitivity (*ts*) was recessive (data not shown). Interestingly, the *ts* phenotype of *sln2* cells could not be rescued by either *UBR1* (Fig. 1 D) or *HIS3* (data not shown), indicating that the *sln2* mutation affected not only histidine uptake. Since the *sln2*/*SLN2* (AVY200/AVY199) diploid sporulated inefficiently, and the viability of the few spores that did form was very low, it was difficult to preclude, through tetrad analysis, the alternative possibility that *sln2* cells actually bore more than one mutant gene. Therefore we attempted to identify a gene(s) that would rescue the *ts* phenotype of *sln2* cells, and thereafter to determine whether that gene could rescue the histidine-uptake defect as well.

A standard complementation screen (see Materials and methods) yielded *CIN8*, a previously described gene encoding a kinesin-like protein (Hoyt et al. 1992; Roof et al. 1992). *CIN8* was found to alleviate not only the *ts* phenotype of *sln2* cells, but their histidine-uptake defect as well (Fig. 6 B), suggesting that the sought *SLN2* gene may be *CIN8*. To determine whether the AVY200 (*sln2* P_{GAL}-*UBR1*) strain carried a mutant allele of *CIN8*, we isolated it from these cells using the method of gap repair (Fig. 4 B and see Materials and methods). Sequencing of this allele, termed *CIN8**, showed the presence of the A → T mutation at position 1237, resulting in a premature stop codon (TGA) at codon 413 of the 1038-codon *CIN8* ORF (data not shown). The resulting truncated product would lack the entire stalk/tail region and a portion of the putative motor domain (Hoyt et al. 1992). As expected, *CIN8**, in contrast to *CIN8*, was unable to rescue the histidine-uptake defect (inviability on SD plates) of AVY200 cells (Fig. 6 B). These results made it very likely that the *SLN2* gene was in fact *CIN8*.

To produce independent evidence bearing on this conjecture, we deleted the bulk of *CIN8* from the JD52 (*his3Δ UBR1*) and JD54 (*his3Δ* P_{GAL}-*UBR1*) strains (the latter strain is Ubr1$^-$ on dextrose) (see Fig. 4 C and Materials and methods). The resulting *cin8Δ* strains AVY203 (*his3Δ cin8Δ UBR1*) and AVY202 (*his3Δ cin8Δ* P_{GAL}-*UBR1*) (Table 1) were tested for viability on SD plates, i.e., for complementation of the histidine uptake defect. Whereas the AVY203 (*cin8Δ UBR1*) cells were viable under these conditions, the AVY202 (*his3Δ cin8Δ* P_{GAL}-*UBR1*) cells were inviable (Fig. 6 C), a set of phenotypes indistinguishable from that conferred by the *sln2* mutation. Moreover, as would have be expected if *SLN2* and *CIN8* were the same gene, the *HIS3* gene was found to rescue Ubr1p-lacking *cin8Δ* cells on SD plates (Fig. 6 D), identically to the effect of *HIS3* on *sln2* cells (Fig. 2 A). Taken together, these results (Figs. 4 and 6) indicated that *sln2* was a defective (truncated) allele of the *CIN8* gene.

Discussion

We report the following results.

(1) A synthetic lethal screen for *S. cerevisiae* mutants that require the presence of the N-end rule pathway (specifically, the presence of its recognition component Ubr1p) for viability yielded a mutant termed *sln2*. The viability of *sln2 ubr1* cells (specifically, of *sln2* P_{GAL}-*UBR1* cells on dextrose-containing media) could be rescued by *UBR1* expressed at either normal or increased levels.

(2) The *HIS3* gene (which was absent from the parental "wild-type" strain) was found to rescue the inviability of Ubr1p-lacking *sln2* cells. High (>75 μg/ml) concentrations of histidine in the medium could also rescue these cells.

(3) This defect in histidine uptake, exhibited by Ubr1p-lacking *sln2* cells but not by *SLN2 ubr1Δ* cells, was traced to the *HIP1* gene, encoding the histidine transporter. The *HIP1* allele in Ubr1p-lacking *sln2* cells was wild type, but the level of *HIP1* mRNA in these cells was reproducibly about 2-fold lower than in parental (*SLN2*) cells. In addition, the relative rate of ^{14}C-histidine uptake by Ubr1p-lacking *sln2 HIS3* cells was about 35% of that by *SLN2 HIS3* cells and about 50% of that by *sln2 UBR1 HIS3* cells. The incomplete shutoff of histidine import in Ubr1p-lacking *sln2* cells was consistent with their rescue by high concentrations of histidine in the medium. Further, the incomplete restoration of histidine import in Ubr1p-expressing *sln2* cells was consistent with their detectably slower growth on SD plates in comparison to either *SLN2 UBR1* or *SLN2 ubr1Δ* cells. Finally, a doubling of the *HIP1* copy number in Ubr1p-lacking *sln2* cells partially complemented their inviability on SD plates. Thus, the histidine-uptake defect of Ubr1p-lacking *sln2* cells was caused at least in part by underexpression of Hip1p.

(4) Yet another property of the *sln2* mutant was its inviability at 37 °C, which could not be rescued by either *HIS3* or *UBR1*. This made possible the cloning of *SLN2*, which was found to be a previously characterized gene called *CIN8*, encoding a kinesin-like protein (Hoyt et al. 1992; Roof et al. 1992). The dependence of histidine transport by *cin8Δ* cells on the presence of Ubr1p is a previously unknown property of *cin8Δ* cells.

Cin8p participates in the assembly of a bipolar spindle. *cin8* mutants are viable, owing to the presence of Kip1p, another kinesin-related motor protein and a functional homolog of Cin8p; at least one of the two proteins is required for cell viability (Hoyt et al. 1992; Roof et al. 1992). Geiser et al. (1997) identified a number of other genes that are required for viability in the absence of *CIN8*. These genes encode proteins that function in cell-division pathways (spindle motors, microtubule stabilizers, checkpoint regulators).

In retrospect, it was an accident of the absence of *HIS3* (rather than of another genetic marker) from the parental JD54 strain (P_{GAL}-*UBR1*) that led us to the finding, through a synthetic lethal screen, of a connection between a kinesin-like protein, the N-end rule pathway, and regulation of the histidine import. Preliminary tests for the ability of Ubr1p-lacking *sln2* cells to import either dipeptides or several amino acids other than histidine indicated approximately wild-type rates of uptake (data not shown).

A short-lived substrate of Ubr1p that mediates its influence on the phenotype of *cin8Δ* (*sln2*) cells is unknown. At the same time, it is clear that a connection between the N-end rule pathway and a Cin8p-dependent process involves the expression of the histidine transporter Hip1p. A parsimonious model which is consistent with the available evidence posits a short-lived transcriptional repressor of the *HIP1* gene that is degraded by the N-end rule pathway, and is also down-regulated, in an unknown manner, through a Cin8p-dependent pathway (Fig. 7 A). This model is analogous to the previously deciphered circuit that controls the expression of the peptide transporter Ptr2p in *S. cerevisiae.* It was found that Cup9p, a homeodomain-containing transcriptional repressor of *PTR2*, is a short-lived protein targeted for Ubr1p-dependent degradation, apparently through an internal degron of Cup9p (Alagramam et al. 1995; Byrd et al. 1998). In the absence of Ubr1p, the increased concentration of Cup9p results in a virtually complete shutoff of the *PTR2* gene and cessation of peptide import (Byrd et al. 1998). We know (see Results) that Cup9p is not involved in the regulation of *HIP1* expression.

An alternative model invokes a positive regulator of *HIP1* that is positively regulated by both the N-end rule and Cin8p-mediated pathways (Fig. 7 B). A transporter homolog Ssy1p of *S. cerevisiae* has been shown to be required for the expression of several amino-acid transporters, including Bap2p, Bap3p and Tat1p, and also the peptide transporter Ptr2p (Didion et al. 1998). Ssy1p appears to function at least in part as a sensor of specific amino acids in the medium. Although *SSY1* is unnecessary for histidine uptake (Didion et al. 1998), it is possible that an analogous plasma membrane-embedded sensor activates the synthesis of other transporters, including Hip1p. The absence of Cin8p may affect the secretory pathway, resulting in lower amounts of the postulated sensor in the plasma membrane. Ubr1p, the E3 component of the ubiquitin-dependent N-end rule pathway, may up-regulate this sensor; for example, through the degradation of another E3 enzyme that targets the sensor for endocytosis and degradation in the vacuole. The involvement of the ubiquitin system in the regulated endocytosis and lysosome/vacuole delivery of transmembrane proteins has previously been demonstrated in both yeast and multicellular eukaryotes (Galan et al. 1996; Hicke and Riezman 1996).

One prediction of the repressor-based model (Fig. 7 A) is that inactivation of the repressor of *HIP1* should suppress the down-regulation of *HIP1* caused by the absence of both Cin8p and Ubr1p. A genetic screen for mutants of this class is under way. An analogous earlier screen for suppressors of the inability of *ubr1Δ* cells to import peptides has led to the identification of Cup9p as a transcriptional repressor of the peptide transporter gene *PTR2* (Byrd et al. 1998).

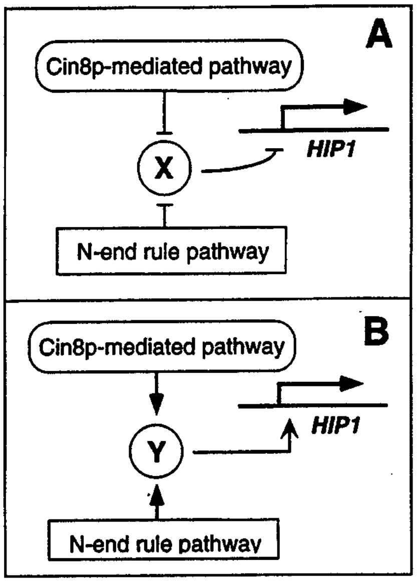


Fig. 7 A,B Possible interconnections between the N-end rule pathway, the regulation of histidine import, and a kinesin-dependent pathway. **A** In this, apparently most-parsimonious model, a short-lived transcriptional repressor of the *HIP1* gene is degraded by the N-end rule pathway, and is *also* down-regulated, in an unknown manner, through a Cin8p-dependent pathway. **B** in an alternative model, a positive regulator of *HIP1* is positively regulated by both the N-end rule and Cin8p-mediated pathways

Acknowledgements We thank members of the Varshavsky laboratory, especially Glenn Turner, Hong-Rui Wang, Fang-Yong Du, Anna Kashina, Hai Rao and Frederic Navarro-Garcia, for helpful discussions and comments on the manuscript. This work was supported by a grant to A.V. from the NIH (DK39520). Y.X. was supported by a postdoctoral fellowship from the NIH.

References

Alagramam K, Naider F, Becker JM (1995) A recognition component of the ubiquitin system is required for peptide transport in *Saccharomyces cerevisiae.* Mol Microbiol 15:225–234

Ausubel FM, Brent R, Kingston RE, Moore DD, Smith JA, Seidman JG, Struhl K (eds) (1996) Current protocols in molecular biology. Wiley-Interscience, New York

Baker RT, Varshavsky A (1995) Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway. J Biol Chem 270:12065–12074

Bachmair A, Finley D, Varshavsky A (1986) In vivo half-life of a protein is a function of its amino-terminal residue. Science 234:179–186

Bachmair A, Varshavsky A (1989) The degradation signal in a short-lived protein. Cell 56:1019–1032

Balzi E, Choder M, Chen W, Varshavsky A, Goffeau A (1990) Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae.* J Biol Chem 265:7464–7471

Bartel B, Wünning I, Varshavsky A (1990) The recognition component of the N-end rule pathway. EMBO J 9:3179–3189

Baudin A, Ozier-Kalogeropolous O, Denoel A, Lacroute A, Cullin C (1993) A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. Nucleic Acids Res 21:3329–3330

Baumeister W, Walz J, Zühl F, Seemüller E (1998) The proteasome: paradigm of a self-compartmentalizing protease. Cell 92:367–380

Byrd C, Turner GC, Varshavsky A (1998) The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. EMBO J 17:269–277

Chau V, Tobias JW, Bachmair A, Marriott D, Ecker DJ, Gonda DK, Varshavsky A (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science 243:1576–1583

Chen D-C, Yang B-C, Kuo T-T (1992) One-step transformation of yeast in stationary phase. Curr Genet 21:83–84

Coux O, Tanaka K, Goldberg AL (1996) Structure and functions of the 20S and 26S proteasomes. Annu Rev Biochem 65:801–817

deGroot RJ, Rümenapf T, Kuhn RJ, Strauss JH (1991) Sindbis virus RNA polymerase is degraded by the N-end rule pathway. Proc Natl Acad Sci USA 88:8967–8971

Didion T, Regenberg B, Jørgesen MU, Kielland-Brandt MC, Andersen HA (1998) The permease homologue Ssy1p controls the expression of amino-acid and peptide transporter genes in *Saccharomyces cerevisiae*. Mol Microbiol 27:643–650

Dohmen RJ, Wu P, Varshavsky A (1994) Heat-inducible degron: a method for constructing temperature-sensitive mutants. Science 263:1273–1276

Galan JM, Moreau V, Andre B, Volland C, Haguenauer-Tsapis R (1996) Ubiquitination mediated by the Npi1p/Rsp5b ubiquitin-protein ligase is required for endocytosis of the yeast uracil permease. J Biol Chem 271:10946–10952

Geiser JR, Schott EJ, Kingsbury TJ, Cole NB, Totis LJ, Bhattacharyya G, He L, Hoyt MA (1997) *Saccharomyces cerevisiae* genes required in the absence of the *CIN8*-encoded spindle motor act in functionally diverse mitotic pathways. Mol Cell Biol 8:1035–1050

Ghislain M, Dohmen RJ, Levy F, Varshavsky A (1996) Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in *Saccharomyces cerevisiae*. EMBO J 15:4884–4899

Glickman MH, Rubin DM, Fried VA, Finley D (1998) The regulatory particle of the *Saccharomyces cerevisiae* proteasome. Mol Cell Biol 18:3149–3162

Gonda DK, Bachmair A, Wünning I, Tobias JW, Lane WS, Varshavsky A (1989) Universality and structure of the N-end rule. J Biol Chem 264:16700–16712

Gonen H, Schwartz AL, Ciechanover A (1991) Purification and characterization of a novel protein that is required for degradation of N-alpha-acetylated proteins by the ubiquitin system. J Biol Chem 266:19221–19231

Grauslund M, Didion T, Kielland-Brandt MC, Andersen HA (1995) *BAP2*, a gene encoding a permease for branched-chain amino acids in *Saccharomyces cerevisiae*. Biochim Biophys Acta 1269:275–280

Hein C, Springael J-Y, Christiane V, Haguenauer-Tsapis R, Andre B (1995) *NPI1*, an essential yeast gene involved in induced degradation of Gap1 and Fur4 permeases, encodes the Ras5 ubiquitin-protein ligase. Mol Microbiol 18:77–87

Hershko A (1991) The ubiquitin pathway for protein degradation. Trends Biochem Sci 16:265–268

Hershko A (1997) Roles of ubiquitin-mediated proteolysis in cell cycle control. Curr Opin Cell Biol 9:788–799

Herskowitz I, Jensen RE (1991) Putting the *HO* gene to work: practical uses for mating-type switching. Methods Enzymol 194:132–146

Hicke L, Riezman H (1996) Ubiquitination of a yeast plasma membrane receptor signals its ligand-stimulated endocytosis. Cell 84:277–287

Hill CP, Johnston NL, Cohen RE (1993) Crystal structure of a ubiquitin-dependent degradation substrate: a three-disulfide form of lysozyme. Proc Natl Acad Sci USA 90:4136–4140

Ho AK, Raczniak GA, Ives EB, Wente SR (1998) The integral membrane protein snl1p is genetically linked to yeast nuclear pore complex function. Mol Biol Cell 9:355–373

Hochstrasser M (1996) Ubiquitin-dependent protein degradation. Annu Rev Genet 30:405–439

Hondermarck H, Sy J, Bradshaw RA, Arfin SM (1992) Dipeptide inhibitors of ubiquitin-mediated protein turnover prevent growth factor-induced neurite outgrowth in rat pheochromocytoma PC12 cells. Biochem Biophys Res Commun 30:280–288

Hoyt MA, He L, Loo KK, Saunders WS (1992) Two *Saccharomyces cerevisiae* kinesin-related gene products required for mitotic spindle assembly. J Cell Biol 118:109–120

Jauniaux J-C, Grenson M (1990) *GAP1*, the general amino-acid permease gene of *Saccharomyces cerevisiae*: nucleotide sequence, protein similarity with the other bakers yeast amino-acid permeases, and nitrogen catabolite repression. Eur J Biochem 190:39–44

Johnson ES, Ma PC, Ota IM, Varshavsky A (1995) A proteolytic pathway that recognizes ubiquitin as a degradation signal. J Biol Chem 270:17442–17456

Kwon YT, Reiss Y, Fried VA, Hershko A, Yoon JK, Gonda DK, Sangan P, Copeland NG, Jenkins NA, Varshavsky A (1998) The mouse and human genes encoding the recognition component of the N-end rule pathway. Proc Natl Acad Sci USA 95:7898–7903

Madura K, Varshavsky A (1994) Degradation of G-alpha by the N-end rule pathway. Science 265:1454–1458

Merrill BJ, Holm C (1998) The RAD52 recombinational repair pathway is essential in pol30 (PCNA) mutants that accumulate small single-stranded DNA fragments during DNA synthesis. Genetics 148:611–624

Mumberg D, Muller R, Funk M (1994) Regulatable promoters of *Saccharomyces cerevisiae* – comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Res 22:5767–5768

Ota IM, Varshavsky A (1993) A yeast protein similar to bacterial two-component regulators. Science 262:566–569

Peters J-M, King RW, Deshaies RJ (1998) Cell cycle control by ubiquitin-dependent proteolysis. Ubiquitin and the biology of the cell. Plenum Press, New York, pp 345–387

Pickart CM (1997) Targeting of substrates to the 26S proteasome. FASEB J 11:1055–1066

Posas F, Wurgler-Murphy SM, Maeda T, Witten EA, Thai TC, Saito H (1996) Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 "two-component" osmosensor. Cell 86:865–875

Rechsteiner M, Hoffman L, Dubiel W (1993) The multicatalytic and 26S proteases. J Biol Chem 268:6065–6068

Roof DM, Meluh PB, Rose MD (1992) Kinesin-related proteins required for assembly of the mitotic spindle. J Cell Biol 118:95–108

Rothstein R (1991) Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol 194:281–301

Scheffner M, Smith S, Jentsch S (1998) The ubiquitin conjugation system. In: Peters J-M, Harris JR, Finley D (eds) Ubiquitin and the biology of the cell. Plenum Press, New York, pp 65–98

Schmitt ME, Brown TA, Trumpower BL (1990) A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. Nucleic Acids Res 18:3091–3092

Sherman F, Fink GR, Hicks JB (1986) Methods in yeast genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Sijts AJ, Pilip I, Pamer EG (1997) The *Listeria* monocytogenes-secreted p60 protein is an N-end rule substrate in the cytosol of infected cells. Implications for major histocompatibility complex class-I antigen processing of bacterial proteins. J Biol Chem 272:19261–19268

Sikorski RS, Hieter P (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *S. cerevisiae*. Genetics 122:19–27

Singer-Kruger B, Ferro-Novick S (1997) Use of a synthetic lethal screen to identify yeast mutants impaired in endocytosis, vacuolar protein sorting and the organization of the cytoskeleton. Eur J Cell Biol 74:365–375

Solomon V, Baracos V, Sarraf P, Goldberg A (1998 a) Rates of ubiquitin conjugation increase when muscles atrophy, largely through activation of the N-end rule pathway. Proc Natl Acad Sci USA 95:12602–12607

Solomon V, Lecker SH, Goldberg AL (1998 b) The N-end rule pathway catalyzes a major fraction of the protein degradation in skeletal muscle. J Biol Chem 273:25216–25222

Taban CH, Hondermarck H, Bradshaw RA, Boilly B (1996) Effect of a dipeptide inhibiting ubiquitin-mediated protein degradation on nerve-dependent limb regeneration in the newt. Experientia 52:865–870

Tanaka J, Fink GR (1985) The histidine permease gene (*HIP1*) of *Saccharomyces cerevisiae*. Gene 38:205–214

Tobias JW, Shrader TE, Rocap G, Varshavsky A (1991) The N-end rule in bacteria. Science 254:1374–1377

Varshavsky A (1996) The N-end rule: functions, mysteries, uses. Proc Natl Acad Sci USA 93:12142–12149

Varshavsky A (1997) The ubiquitin system. Trends Biochem Sci 22:383–387

Varshavsky A, Byrd C, Davydov IV, Dohmen RJ, Du F, Ghislain M, Gonzalez M, Grigoryev S, Johnson ES, Johnsson N, Johnston JA, Kwon YT, Lévy F, Lomovskaya O, Madura K, Ota I, Rümenapf T, Shrader TE, Suzuki T, Turner G, Waller PRH, Webster A (1998) The N-end rule pathway. In: Peters J-M, Harris JR, Finley D (eds) Ubiquitin and the biology of the cell. Plenum Press, New York, pp 223–278

Worley CK, Ling R, Callis J (1998) Engineering in vivo instability of firefly luciferase and *Escherichia coli* beta-glucuronidase in higher plants using recognition elements from the ubiquitin pathway. Plant Mol Biol 37:337–347

MOLECULAR AND CELLULAR BIOLOGY, Jan. 1999, p. 182–193 Vol. 19, No. 1
0270-7306/99/$04.00+0
Copyright © 1999, American Society for Microbiology. All Rights Reserved.

Alternative Splicing Results in Differential Expression, Activity, and Localization of the Two Forms of Arginyl-tRNA-Protein Transferase, a Component of the N-End Rule Pathway

YONG TAE KWON, ANNA S. KASHINA, AND ALEXANDER VARSHAVSKY*

Division of Biology, California Institute of Technology, Pasadena, California 91125

Received 13 August 1998/Returned for modification 21 September 1998/Accepted 6 October 1998

The N-end rule relates the in vivo half-life of a protein to the identity of its N-terminal residue. The underlying ubiquitin-dependent proteolytic system, called the N-end rule pathway, is organized hierarchically: N-terminal aspartate and glutamate (and also cysteine in metazoans) are secondary destabilizing residues, in that they function through their conjugation, by arginyl-tRNA-protein transferase (R-transferase), to arginine, a primary destabilizing residue. We isolated cDNA encoding the 516-residue mouse R-transferase, ATE1p, and found two species, termed *Ate1-1* and *Ate1-2*. The *Ate1* mRNAs are produced through a most unusual alternative splicing that retains one or the other of the two homologous 129-bp exons, which are adjacent in the mouse *Ate1* gene. Human *ATE1* also contains the alternative 129-bp exons, whereas the plant (*Arabidopsis thaliana*) and fly (*Drosophila melanogaster*) *Ate1* genes encode a single form of ATE1p. A fusion of ATE1-1p with green fluorescent protein (GFP) is present in both the nucleus and the cytosol, whereas ATE1-2p–GFP is exclusively cytosolic. Mouse ATE1-1p and ATE1-2p were examined by expressing them in *ate1Δ Saccharomyces cerevisiae* in the presence of test substrates that included Asp-βgal (β-galactosidase) and Cys-βgal. Both forms of the mouse R-transferase conferred instability on Asp-βgal (but not on Cys-βgal) through the arginylation of its N-terminal Asp, the ATE1-1p enzyme being more active than ATE1-2p. The ratio of *Ate1-1* to *Ate1-2* mRNA varies greatly among the mouse tissues; it is ~0.1 in the skeletal muscle, ~0.25 in the spleen, ~3.3 in the liver and brain, and ~10 in the testis, suggesting that the two R-transferases are functionally distinct.

The half-lives of intracellular proteins range from a few seconds to many days. The rates of processive proteolysis are a function of the cell's physiological state and are controlled differentially for specific proteins. In particular, most of the damaged or otherwise abnormal proteins are metabolically unstable. Many other proteins, while long-lived as components of larger macromolecular structures such as ribosomes and oligomeric proteins, are metabolically unstable as free subunits. Regulatory proteins are often also short-lived in vivo, providing a way to generate their spatial gradients and to rapidly adjust their concentrations, or subunit compositions, through changes in the rate of their synthesis or degradation (20, 23, 28, 39, 44, 55).

The posttranslational conjugation of arginine (Arg) to the N termini of eukaryotic proteins was described 35 years ago (26), but the function of this modification, and of the enzyme involved, Arg-tRNA-protein transferase (R-transferase) (47), remained unknown until the discovery that the identity of N-terminal residue in a protein influences its metabolic stability (4). The resulting relation was termed the N-end rule (54). Aspartate (Asp) and glutamate (Glu), the two N-terminal residues known to be arginylated by R-transferase (47), were shown to be destabilizing residues in the N-end rule (4). It was therefore proposed (4) that the function of R-transferase is to target proteins for degradation by conjugating Arg, one of the primary destabilizing residues, to secondary destabilizing N-terminal residues (Asp and Glu in fungi; Asp, Glu, and Cys in metazoans (18) (Fig. 1). It was also proposed (4) that the analogous prokaryotic enzyme Leu, Phe-tRNA-protein transferase (L, F-transferase) (47) mediates the activity of N-terminal Arg and Lys, which, in prokaryotes, would be the secondary destabilizing residues. These conjectures were confirmed (7, 17, 45, 53).

The similar but distinct degradation signals which together give rise to the N-end rule are called the N-degrons (54, 56). In eukaryotes, an N-degron comprises two determinants: a destabilizing N-terminal residue and an internal Lys residue of a substrate (5, 22). The Lys residue is the site of formation of a substrate-linked multiubiquitin chain (11). The N-end rule pathway is thus one pathway of the ubiquitin (Ub) system. Ub is a 76-residue protein whose covalent conjugation to other proteins plays a role in a multitude of processes, including cell growth, division, differentiation, and responses to stress (20, 23, 39, 55). In many of these processes, Ub acts through routes that involve the degradation of Ub-protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease (9, 13, 40, 43).

(In the text that follows, names of mouse genes are in italics, with the first letter uppercase. Names of human and *Saccharomyces cerevisiae* genes are also in italics, all uppercase. If human and mouse genes are named in the same sentence, the mouse gene notation is used. Names of *S. cerevisiae* proteins are roman, with the first letter uppercase and an extra lowercase "p" at the end. Names of the corresponding mouse and human proteins are the same, except that all letters but the last "p" are uppercase. The latter usage is a modification of the existing convention (50), to facilitate simultaneous discussions of yeast, mouse, and human proteins. In some citations, the abbreviated name of a species precedes the gene's name.)

The N-end rule is organized hierarchically. In the yeast *S. cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by the *NTA1*-encoded N-terminal amidase (Nt-amidase) (6), to the secondary destabilizing N-terminal residues Asp and Glu. The

---

* Corresponding author. Mailing address: Division of Biology, 147-75, Caltech, 1200 East California Blvd., Pasadena, CA 91125. Phone: (626) 395-3785. Fax: (626) 440-9821. E-mail: avarsh@cco.caltech.edu.

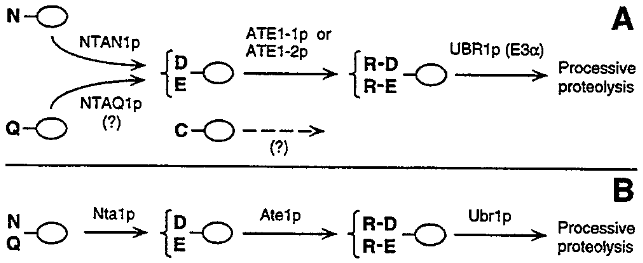


FIG. 1. Comparison of enzymatic reactions that underlie the activity of the tertiary and secondary destabilizing residues among eukaryotes. (A) Mammals (reference 54 and this work); (B) the yeast *S. cerevisiae* (5). N-terminal residues are indicated by single-letter abbreviations for amino acids; ovals denote the rest of a protein substrate. The *Ntan1*-encoded mammalian Nt-amidase converts N-terminal Asn to Asp, whereas N-terminal Gln is deamidated by a distinct Nt-amidase that remains to be identified (19, 51). In contrast, the yeast Nt-amidase can deamidate either N-terminal Asn or Gln (6). The secondary destabilizing residues Asp and Glu are arginylated by the mammalian ATE1-1p or ATE1-2p R-transferase (see Results). A Cys-specific mammalian R-transferase remains to be identified (see Results). N-terminal Arg, one of the primary destabilizing residues (54), is recognized by N-recognin (E3) (see the introduction).

destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *S. cerevisiae ATE1*-encoded Arg-tRNA-protein transferase (R-transferase), to Arg, one of the primary destabilizing residues (7) (Fig. 1B). In mammals, the deamidation step is bifurcated, in that two distinct Nt-amidases specific, respectively, for N-terminal Asn and Gln, mediate the activity of tertiary destabilizing residues (19, 51) (Fig. 1A). Mice lacking the Asn-specific Nt-amidase NTAN1p have recently been produced through targeted mutagenesis and found to be fertile, outwardly normal, but behaviorally distinct from their congenic wild-type counterparts (28a). In mammals, the set of secondary destabilizing residues contains not only Asp and Glu but also Cys, which is a stabilizing residue in yeast (18, 54) (Fig. 1).

The primary destabilizing N-terminal residues are bound directly by the *UBR1*-encoded N-recognin (also called E3α), the recognition component of the N-end rule pathway (8). In *S. cerevisiae*, N-recognin is a 225-kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues, Phe, Leu, Trp, Tyr, Ile, Arg, Lys, and His (54). The *Ubr1* genes encoding mouse and human N-recognins, also called Eα (21, 41), have been cloned (29), and mouse strains lacking *Ubr1* have recently been constructed (29a).

The known functions of the N-end rule pathway include the control of peptide import in *S. cerevisiae*, through the degradation of Cup9p, a transcriptional repressor of *PTR2* which encodes the peptide transporter (2, 10); a role in regulating the Sln1p-dependent phosphorylation cascade that mediates osmoregulation in *S. cerevisiae* (38); the degradation of Gpa1p, a Gα protein of *S. cerevisiae* (34); and the conditional degradation of alphaviral RNA polymerase in virus-infected metazoan cells (16, 54). Physiological N-end rule substrates were also identified among the proteins secreted into the host cell's cytosol by intracellular parasites such as the bacterium *Listeria monocytogenes* (46). Inhibition of the N-end rule pathway was reported to interfere with mammalian cell differentiation (24) and to delay limb regeneration in amphibians (52). Microarray-based comparisons of gene expression patterns in wild-type and congenic *ubr1*Δ strains of *S. cerevisiae* have shown that a number of yeast genes, of diverse functions, are significantly up- or down-regulated in the absence of the N-end rule pathway (39a).

The mammalian counterpart of the yeast *ATE1*-encoded R-transferase was partially purified from rabbit reticulocytes and shown to cofractionate with Arg-tRNA synthetase (12). Recent studies of the Ub-dependent proteolysis of endogenous proteins in muscle extracts suggested that the N-end rule pathway plays a major role in catabolic states that result in muscle atrophy (48, 49). A significant fraction of the N-end rule pathway's activity in muscle extracts was found to be tRNA dependent, indicating the involvement of R-transferase (48, 49). It was also reported that a crush injury to the rat sciatic nerve results in a ~10-fold increase in the rate of arginine conjugation to the N termini of unidentified proteins in the nerve's region upstream of the crush site (15, 57), suggesting an injury-induced increase in the concentration of R-transferase substrates and/or an enhanced activity of the N-end rule pathway.

In this work, we began the functional analysis of mammalian R-transferase (ATE1p) by isolating mouse cDNA encoding this enzyme. Surprisingly, we found two *Ate1* cDNA species, which were identical except for a 129-bp region that encoded similar but distinct sequences. One or the other, but not both, of the corresponding *Ate1* exons is retained in the mature *Ate1* mRNA, and the ratio of the resulting two species, *Ate1-1* and *Ate1-2*, varies greatly among mouse tissues. We also show that ATE1-1p and ATE1-2p, while differing in activity, can arginylate N-terminal Asp and Glu in model substrates. However, neither of them can arginylate N-terminal Cys, the known secondary destabilizing residue (18), suggesting the existence of a distinct Cys-specific mammalian R-transferase.

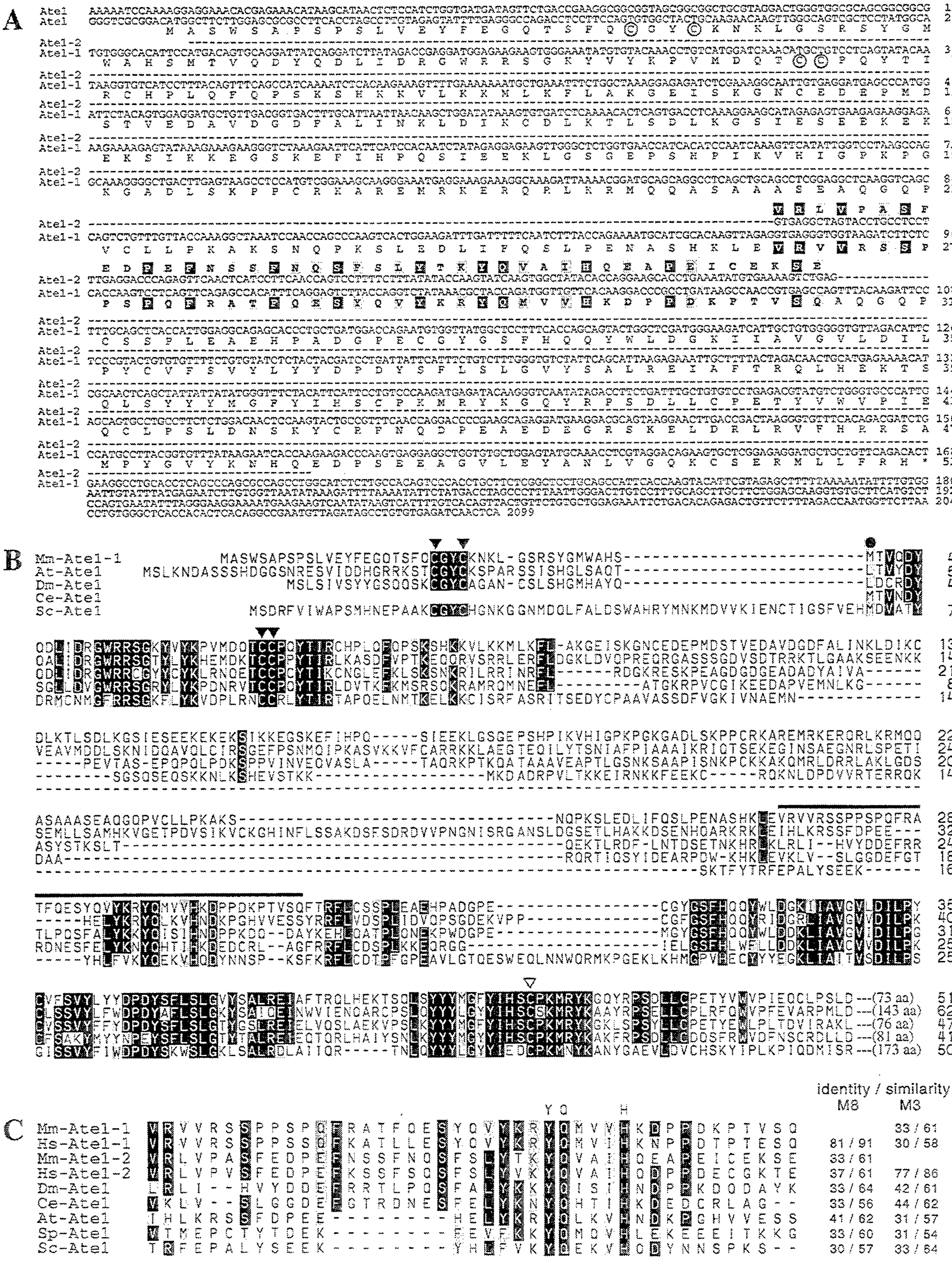
MATERIALS AND METHODS

Strains and plasmids. The *S. cerevisiae* strains used were JD55 (*MATa ura3-52 his3-Δ200 leu2-3,112 trp1-Δ63 lys2-801 ubr1Δ::HIS3*) (34) and SGY3 (*MATα ura3-52 lys2-801 ade2-101 trp1-Δ63 his3-Δ200 leu2-Δ1ate1-Δ2::LEU2*) (19a). Cells were grown in rich medium (YPD) or in synthetic media (SD) containing 0.67% yeast nitrogen base without amino acids (Difco), auxotrophic nutrients, and 2% glucose. To induce the P_{GAL} promoter, glucose was replaced by 2% galactose (SG medium). Transformation of *S. cerevisiae* was performed by the lithium acetate method (3).

The *ubr1*Δ *ate1*Δ double mutant AVY34 was constructed by replacing 93% of the *ATE1* open reading frame (ORF) (the first 470 codons) in strain JD55 (*ubr1*Δ) by the *LEU2* gene, through homologous recombination (42) with the introduced *LEU2* gene flanked on either side by 40 bp of *ATE1*-specific sequences. Mutants were selected on SD (lacking Leu and His) plates, and Leu$^+$ isolates were checked by PCR for the absence of *ATE1* and by colony assays (7, 8) for the absence of Ate1p and Ubr1p activity. High-copy-number pUB23-X plasmids expressing Ub-X-βgal proteins (see below) from P_{GAL} in *S. cerevisiae* have been described elsewhere (4). Mouse *Ate1-1* and *Ate1-2* cDNAs (see below) were subcloned into the low-copy-number vector p414GAL1 (36), using the engineered *Bam*HI (5′) and *Xho*I (3′) restriction sites, yielding plasmids pAT1 and pAT2. For localization assays with green fluorescent protein (GFP), cDNAs encoding mouse ATE1-1p or ATE1-2p were subcloned into the pEGFP-N1 N-terminal protein fusion vector (Clontech, Palo Alto, Calif.), using the engineered *Xho*I (5′) and *Age*I (3′) restriction sites, yielding plasmids pAT1-GFP and pAT2-GFP.

Isolation of the mouse *Ate1-1* and *Ate1-2* cDNAs. The 392-bp fragment of the mouse EST (expressed sequence tag) clone (accession no. AA415294), which was identified in GenBank through species walking (see Results), was used as a probe to screen a λgt10-based mouse cDNA library from MEL-C19 cells (Clontech), using standard procedures (3). Eight positive clones, whose inserts ranged from 0.5 to 1.6 kb, were analyzed by PCR and partial sequencing. The cDNA inserts of clones 3 and 8 were then subcloned into pBluescript II SK$^+$ (29) and sequenced on both strands. The resulting ORFs were identical except for a 129-bp internal region (see Results) (Fig. 2A). The deduced amino acid sequences of the mouse cDNA clones 3 and 8 were weakly but significantly similar to the deduced sequences of *Caenorhabditis elegans* and *S. cerevisiae* ATE1p (see Results) (Fig. 2B) and corresponded to nucleotides (nt) 699 to 1870 and 587 to 2099, respectively, in the subsequently produced full-length mouse *Ate1-1* and *Ate1-2* cDNAs (accession no. AF079096 and AF079097).

A human EST clone (accession no. AA503372) whose deduced amino acid sequence was highly similar to that of the partial mouse *Ate1* cDNA clone 8 was found in GenBank, using the partial mouse ATE1p sequence as a query. This EST clone was purchased from Genome Systems (St. Louis, Mo.) and sequenced


A
Ate1
Ate1-1
Ate1-2
B
Mm-Ate1-1
At-Ate1
Dm-Ate1
Ce-Ate1
Sc-Ate1
C
identity / similarity
M8
M3
Mm-Ate1-1 33 / 61
Hs-Ate1-1 81 / 91 30 / 58
Mm-Ate1-2 33 / 61
Hs-Ate1-2 37 / 61 77 / 86
Dm-Ate1 33 / 64 42 / 61
Ce-Ate1 33 / 56 44 / 62
At-Ate1 41 / 62 31 / 57
Sp-Ate1 33 / 60 31 / 54
Sc-Ate1 30 / 57 33 / 54

to obtain more of the 5′-proximal human *ATE1* sequence. Reverse transcription-PCR (RT-PCR) (3) was then carried out with poly(A)$^+$ RNA from mouse embryonic fibroblasts, using the mouse *Ate1*-specific reverse primer 5′-CCTTT GGTAACAAACAGACTGGCTG-3′ and the forward primer 5′-TCTCATAG ACCGAGGATGGCGAAG-3′, whose sequence was derived from the above human EST clone. The resulting PCR products, which appeared as a smear upon agarose gel electrophoresis, were ligated into the TA cloning vector (Invitrogen, San Diego, Calif.), and the ligation mixture was used as a template for PCR using a nested mouse *Ate1* primer 5′-CTGCAGCTGAGGCCTGCTGCATCCG-3′ and a vector-specific primer 5′-GTTTTCCCAGTCACGAC-3′. This strategy yielded a single major DNA species (data not shown). We then applied 5′-RACE (rapid amplification of 5′ cDNA ends) (3), using the above RT-PCR-derived sequence, to produce the full-length *Ate1* cDNA as previously described (19).

Analysis of the mouse *Ate1* gene. The mouse genomic DNA from L cells was used as a template for PCR, using the Expand high-fidelity PCR system (Boehringer, Indianapolis, Ind.) and exon-specific primers as previously described (29), to produce DNA fragments that together spanned ~4 kb of the mouse *Ate1* gene and contained the two alternative 129-bp *Ate1* exons. The regions encompassing exon/intron junctions were sequenced by using exon- and intron-specific primers. Thereafter, a strategy described earlier for the *Ubr1* gene (29) was used to screen, using a fragment of the mouse *Ate1* cDNA (nt 255-1139), a BAC (bacterial artificial chromosome)-based library of mouse genomic DNA fragments from strain 129SvJ (Genome Systems), yielding one BAC clone containing the mouse *Ate1* gene.

Isolation of the human, plant, and fly *ATE1* cDNAs. Using the cloned mouse ATE1p sequences (see above) as queries, we identified in GenBank several significantly similar EST sequences from other organisms (data not shown). To determine whether these species also contained the two forms of *ATE1* mRNA, we isolated the corresponding *ATE1* cDNAs. RT-PCR (3) with poly(A)$^+$ RNA from human 293 cells and the primers 5′-CAATGGCATGTGGGCACATTCC ATG-3′ (specific for the human EST clone AA503372 [see above]) and 5′-CC ACAGGTACTGAATATGTATCCTG-3′ (specific for the human EST clone AA195361) was carried out, yielding a 1.6-kb human *ATE1* cDNA fragment lacking the first 41 codons of the *ATE1* ORF. This fragment (a mixture of the two alternative cDNAs) was subcloned into the TA vector (Invitrogen) and sequenced on both strands. Full-length *Ate1* cDNAs from *Arabidopsis thaliana* and *Drosophila melanogaster* were isolated by RT-PCR as well, using total RNA from *A. thaliana* leaves, poly(A)$^+$ RNA from *D. melanogaster* embryos, and primers specific for the 5′ and 3′ ends of the corresponding ORFs. By using the strategy described above for the human *ATE1* cDNAs, the sequences of these primers were derived from the EST clones that were initially identified in GenBank through their similarity to the mouse ATE1p sequence, then purchased from Genome Systems, and sequenced prior to RT-PCR with the corresponding RNA preparations. The final human, plant, and fly *ATE1* cDNAs were sequenced on both strands.

Assays of β-gal. Colony assays for the *Escherichia coli* β-galactosidase (βgal) in *S. cerevisiae* were carried out by overlaying yeast colonies on SG plates with 0.5% agarose containing 0.1% sodium dodecyl sulfate (SDS), 4% dimethylformamide, and a 0.1-mg/ml solution of the chromogenic βgal substrate X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; Calbiochem, La Jolla, Calif.), followed by incubation for 1 to 2 h at 37°C. Quantitative assays for βgal in *S. cerevisiae* were carried out with whole-cell extracts, using another chromogenic βgal substrate, *o*-nitrophenyl-β-D-galactopyranoside (ONPG). Cells in a 5-ml culture (A_{600} of ~1) were pelleted by centrifugation and resuspended in 5 ml of buffer Z (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol [pH 7.0]). After the A_{600} of the suspension was determined 50- or 100-μl samples were diluted to 1 ml with buffer Z; 0.1% SDS (20 μl) and $CHCl_3$ (50 μl) were then added; the suspension was vortexed for 10 to 15 s and incubated for 15 min at 30°C, followed by the addition of 200 μl of ONPG (4 mg/ml in buffer Z) and further incubation at 30°C, until a medium yellow color had developed, at which point the reaction was stopped by the addition of 1 M Na_2CO_3 (0.4 ml). The mixture was centrifuged for 5 min at 1,100 × *g*, and the A_{420} and A_{500} of the samples were measured. The ONPG units (U_{ONPG}) of βgal activity were calculated as follows: $U_{ONPG} = 1{,}000 \times [(A_{420}) - (1.75 \times A_{500})]/t \times v \times A_{600}$, where t and v were, respectively, the time of incubation (minutes) and the sample volume (milliliters) (3).

Purification and N-terminal sequencing of X-βgal proteins. Extracts were prepared (using the liquid nitrogen procedure [3]) from *S. cerevisiae* AVY34 (*ubr1*Δ *ate1*Δ) cotransformed with a pUB23-X plasmid (4) (expressing Ub-X-βgal) and either pAT1 (expressing mouse ATE1-1p) or pAT2 (expressing mouse ATE1-2p). Cultures were grown in SG to an A_{600} of ~1. Specific X-βgal proteins (X = Asp or Cys) were purified by affinity chromatography on ProtoSorb lacZ (Promega, Madison, Wis.), a monoclonal anti-βgal antibody coupled to agarose beads (Promega). X-βgal proteins were further purified by electrophoresis on SDS–7% polyacrylamide gels and were electroblotted onto Immobilon-P^{SQ} membranes (Millipore, Bedford, Mass.). N-terminal sequencing of 10 to 15 pmol of electroblotted X-βgal was carried out for at least five cycles, using an Applied Biosystems 476A protein sequencer (Caltech Microchemistry Facility).

Mouse cell cultures, transfection, and GFP localization. NIH 3T3 cells (ATCC 1658-CRL) were grown as monolayers in Dulbecco's modified Eagle medium (GIBCO, Frederick, Md.) supplemented with 10% fetal bovine serum. Cells for GFP localization analyses were grown to ~15% confluence on glass coverslips for 24 h prior to transfection with either pAT1-GFP or pAT2-GFP, using Lipofectamine (GIBCO) and the manufacturer-supplied protocol. Cells were incubated for 5 h at 37°C in serum-free medium containing DNA and Lipofectamine. Thereafter an equal volume of medium containing 20% serum was added, and the cells were grown for another 12 to 20 h at 37°C. Cells were fixed with 2% formaldehyde in phosphate-buffered saline, and GFP fluorescence was visualized in a Zeiss Axiophot microscope.

Northern hybridization. Mouse multiple-tissue Northern blots containing 2 μg of poly(A)$^+$ RNA per lane (Clontech) were probed with the ^{32}P-labeled 1.1-kb mouse *Ate1* cDNA (nt 638 to 1734), using the manufacturer-supplied protocol.

Determination of the relative levels of *Ate1-1* and *Ate1-2* mRNAs. Samples of total RNA isolated as described previously (3) from mouse spleen, skeletal muscle, liver, brain, testis, and embryonic fibroblasts were subjected to RT-PCR (28 cycles). The primers 5′-CAGTGGAGGATGCTGTTGACGGTGAC-3′ and 5′-GTGCTCTGCCTCCAATGGTGAGCTG-3′ were specific for the identical regions of *Ate1-1* and *Ate1-2* cDNAs that flanked the two 129-bp exons (see Results) which distinguished these cDNAs. The resulting 624-bp product (a mixture of the *Ate1-1* and *Ate1-2* cDNA fragments) was treated with *Scr*FI, which cuts at different sites within the two 129-bp exons, followed by a 2% agarose gel electrophoresis. This procedure made it possible to distinguish the *Ate1-1* and *Ate1-2* fragments. The ratios of the two forms of *Ate1* cDNA were determined by serial dilutions of the samples prior to gel electrophoresis.

Nucleotide sequence accession numbers. The nucleotide sequences reported in this paper were submitted to the GenBank/EMBL data bank and assigned accession no. AF079096 (mouse *Ate1-1* cDNA), AF079097 (mouse *Ate1-2* cDNA), AF079098 (human *Ate1-1* cDNA), AF079099 (human *ATE1-2* cDNA), AF079100 (*A. thaliana Ate1* cDNA), and AF079101 (*D. melanogaster Ate1* cDNA).

RESULTS

Identification of mouse *Ate1* cDNAs by species walking. On the assumption that the sequences of R-transferases in different species might be sufficiently conserved to be detected by using the sequence of the only cloned R-transferase, *S. cerevisiae* Ate1p (7), we have been searching GenBank and related databases. No mammalian sequences in GenBank, including the EST sequences, had significant similarities to *S. cerevisiae* Ate1p. However, we did identify a nematode (*C. elegans*) ORF (accession no. Z21146) that exhibited similarity to yeast Ate1p (Fig. 2B) and then used the *C. elegans* sequence to identify a

FIG. 2. Two forms of the mouse *Ate1* cDNA and the ATE protein family. (A) The mouse *Ate1-1* and *Ate1-2* cDNAs and their products. The nucleotide sequences of *Ate1-2* identical to those of *Ate1-1* (everywhere except for the 129-bp region) are indicated by dashes. In the region of the alternative 129-bp exons of *Ate1-1* and *Ate1-2*, white-on-black and gray shadings highlight, respectively, identical and similar residues. The circled Cys residues are homologous to those that are important for the enzymatic activity of *S. cerevisiae* Ate1p (32). (B) The ATE protein family and the origins of the alternative 129-bp exons. Alignment of the sequences of mouse ATE1-1p (Mm-Ate1-1), *A. thaliana* Ate1p (At-Ate1), *D. melanogaster* Ate1p (Dm-Ate1), *C. elegans* Ate1p (Ce-Ate1), and *S. cerevisiae* Ate1p (Sc-Ate1) (accession no. J05404). Similar residues (gray) were grouped as follows: M, L, I, and V; D, E, N, and Q; R, K, and H; Y, F, and W; S, A, and T. The region encoded by the alternative 129-bp exons of mouse *Ate1* is highlighted by a thick line. Of the Cys residues that are conserved among all ATE proteins, the ones required and not required for the enzymatic activity of *S. cerevisiae* Ate1p (32) are indicated, respectively, by ▼ and ▽. The N-terminally truncated mouse ATE1-1p and ATE1-2p proteins which began at Met-42 (●) lacked the R-transferase activity (data not shown). The highly variable C-terminal regions of ATE proteins were omitted from the alignment. The sequences were aligned using PileUp program (Wisconsin Package; Genetics Computer Group, Madison, Wis.). Gaps (−) were introduced to optimize the alignment. The residue numbers are on the right of the sequences. The sequence of *C. elegans* Ate1p appears to lack the N-terminal region of other ATE proteins because of an error in defining the *Ate1* ORF in the genomic DNA sequence (accession no. Z21146). (C) Alignment of the 43-residue regions that are encoded by the alternative 129-bp exons in mammalian *Ate1*. Sequences shown: mouse (Mm-Ate1-1 and Mm-Ate1-2), human (Hs-Ate1-1 and Hs-Ate1-2), *D. melanogaster* (Dm-Ate1), *C. elegans* (Ce-Ate1), *A. thaliana* (At-Ate1), *S. pombe* (Sp-Ate1; accession no. Z99568), and *S. cerevisiae* (Sc-Ate1) (accession no. J05404). The degrees of identity and similarity of ATE proteins to the deduced amino acid sequences of the M8 (mouse ATE1-1p) or M3 (mouse ATE1-2p) exon of the mouse *Ate1* gene are indicated on the right. The residues conserved among all of the compared sequences are indicated above the alignment.

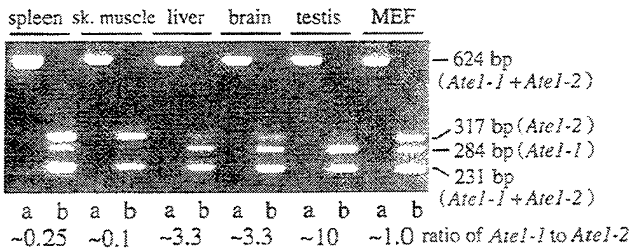

spleen
sk. muscle
liver
brain
testis
MEF
624 bp (*Atel-1* + *Atel-2*)
317 bp (*Atel-2*)
284 bp (*Atel-1*)
231 bp (*Atel-1* + *Atel-2*)
a b a b a b a b a b a b
~0.25 ~0.1 ~3.3 ~3.3 ~10 ~1.0 ratio of *Atel-1* to *Atel-2*

FIG. 3. Expression of *Atel-1* and *Atel-2* mRNAs in different mouse tissues and mouse embryonic fibroblasts (MEF). Two forms of *Atel* cDNA (Fig. 2A) were amplified in a single reaction by RT-PCR, using the same primers, to yield 624-bp fragments that included the region of the alternative 129-bp exons (see Materials and Methods). The 624-bp fragments were digested with *Scr*FI, which produced a 231-bp fragment (a mixture of *Atel-1* and *Atel-2*), a 284-bp, *Atel-1*-specific fragment, and a 317-bp, *Atel-2*-specific fragment. The untreated (lanes a) and *Scr*FI-treated (lanes b) samples from different mouse tissues were analyzed by electrophoresis in a 2% agarose gel. The ratio of the two forms of *Atel* mRNA, defined as the ratio of the 284-bp (*Atel-1*) fragment to the 317-bp (*Atel-2*) fragment, was determined by analyzing serially diluted samples and comparing the resulting band intensities (data not shown). sk., skeletal.

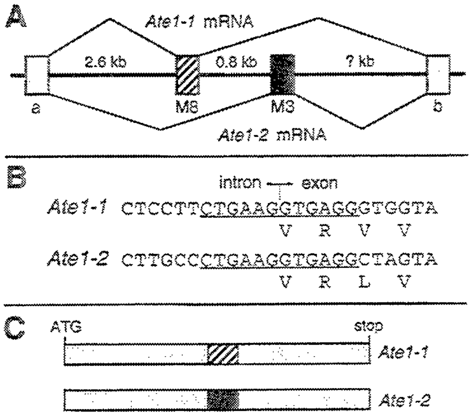


FIG. 4. The two forms of mouse *Atel* mRNA are produced by alternative splicing. (A) The two alternative 129-bp exons are adjacent in the mouse *Atel* gene. The thick line denotes genomic DNA; the striped and black rectangles denote the alternative 129-bp exons, M8 and M3 (see Materials and Methods); gray rectangles denote the flanking *Atel* exons, of unknown sizes; thin lines denote the alternative splicing patterns that yield the two forms of *Atel* mRNA. (B) The underlined 12 bp (6 bp in the intron and 6 bp in the exon) around the splice acceptor sites are identical between the two alternative 129-bp exons. (C) Scale diagrams of the two forms of mouse *Atel* cDNAs. The alternative 129-bp exons M8 and M3 are indicated by the striped and black boxes, respectively.

significantly similar EST sequence of *D. melanogaster* cDNA (accession no. AA391570). Finally, using the deduced amino acid sequence of the *Drosophila* EST clone as a probe, we identified a 467-bp mouse EST sequence (accession no. AA415294) that exhibited weak but significant similarity to the *Drosophila* sequence but no detectable similarity to *S. cerevisiae* Ate1p. On a chance that this 467-bp EST had been derived from the mouse *Atel* cDNA, we used it to screen a mouse cDNA library and indeed isolated the putative mouse *Atel* cDNAs (Fig. 2A).

Alternative splicing results in two species of mouse *Atel* cDNA containing distinct but homologous exons. During the initial mouse cDNA library screening, we found that the cDNA clone 3 (nt 699 to 1870 of *Atel-2* cDNA) was identical to the cDNA clone 8 (nt 587 to 2099 of *Atel-1* cDNA), except for a 129-bp region whose deduced amino acid sequences were similar (31% identity; 61% similarity) (Fig. 2). The two full-length *Atel* cDNAs (termed *Atel-1* and *Atel-2*), which were obtained by RT-PCR followed by 5′-RACE (see Materials and Methods), encoded proteins of identical length, 516 residues (59.2 kDa and pI of 8.14 versus 59.1 kDa and pI of 7.22) (Fig. 2A), that contained regions of similarity to the 57.8-kDa Ate1p of *S. cerevisiae* (Fig. 2B). RT-PCR (followed by subcloning) with RNAs from different mouse tissues also produced the two forms of *Atel* cDNAs, indicating that the two species were in fact present in the initial RNA preparation (Fig. 3 and data not shown).

To determine whether both of the two 129-bp regions of the *Atel-1* and *Atel-2* cDNAs were a part of the *Atel* gene, and whether *Atel-1* and *Atel-2* were produced through alternative splicing, we analyzed the mouse *Atel* gene in the vicinity of its two 129-bp exons, using at first PCR and subsequently a BAC clone containing *Atel* (see Materials and Methods). The two 129-bp exons were located next to each other in the *Atel* gene (Fig. 4A). We also found that the 12-bp sequences around the splice acceptor sites of these exons (6 bp in the intron and 6 bp in the exon) were identical between the two exons (Fig. 4B), consistent with the alternative presence of these exons in the mature *Atel* mRNA. The exon-containing RT-PCR products from different mouse tissues appeared as a single major band retaining one of the two 129-bp exons (Fig. 3 and data not shown). Subcloning and analyses of these RT-PCR products yielded no other differentially spliced *Atel* cDNAs (for example, cDNAs retaining both or neither of the two 129-bp exons), suggesting that the splicing of *Atel* pre-mRNA is tightly regulated to retain one and only one of the two alternative exons. Thus, the two forms of *Atel* mRNAs are produced by a nearly unprecedented (see Discussion) splicing pathway which ultimately yields two proteins of identical size that bear two alternative, homologous but distinct 43-residue internal sequences (Fig. 4C).

The absence of alternative 129-bp exons from the plant and fly *Atel* genes. To explore the evolution of *Atel*, and especially the phylogeny of its alternative 129-bp exons, we cloned the human, plant (*A. thaliana*), and fly (*D. melanogaster*) *ATE1* cDNAs (see Materials and Methods). Two forms of the human *ATE1* cDNA, termed *Hs-ATE1-1* and *Hs-ATE1-2*, were isolated from human 293 cells (the forms' molar ratio was about 1). However, only one form of the *Atel* cDNA was isolated from either the leaves of *A. thaliana* (termed *At-Atel*) or *D. melanogaster* embryos (termed *Dm-Atel*), suggesting that the alternative 129-bp exons may not be present in the *Atel* genes of plants and arthropods. The *A. thaliana* and *D. melanogaster* Ate1p proteins were, respectively, 629 and 477 residues long (71 and 55 kDa, with pIs of 6.0 and 8.4). Mouse ATE1-1p was 82, 38, and 42% identical (as well as 91, 57, and 61% similar) to human ATE1-1p, *A. thaliana* Ate1p, and *D. melanogaster* Ate1p, respectively (Fig. 2 and data not shown). *A. thaliana* Ate1p bore a 16-residue region containing exclusively Asp or Glu (data not shown).

We used RT-PCR and RNA preparations from *A. thaliana* and *D. melanogaster* to amplify the relevant regions of the corresponding *Atel* cDNAs. The resulting fragments were digested with restriction enzymes that recognize, in each species, exclusively the region that corresponds to the 129-bp exons of the mouse *Atel* cDNAs, and the products were analyzed by gel electrophoresis. The initial cDNA fragments of *A. thaliana* and *D. melanogaster Atel* completely disappeared after this treatment, in contrast to the homologous mouse cDNA fragment (which contained two distinct sequences of identical length), suggesting that the two alternative exons were absent from the *Atel* genes of plants and arthropods (Fig. 5A).

While this analysis was under way, complete sequences of

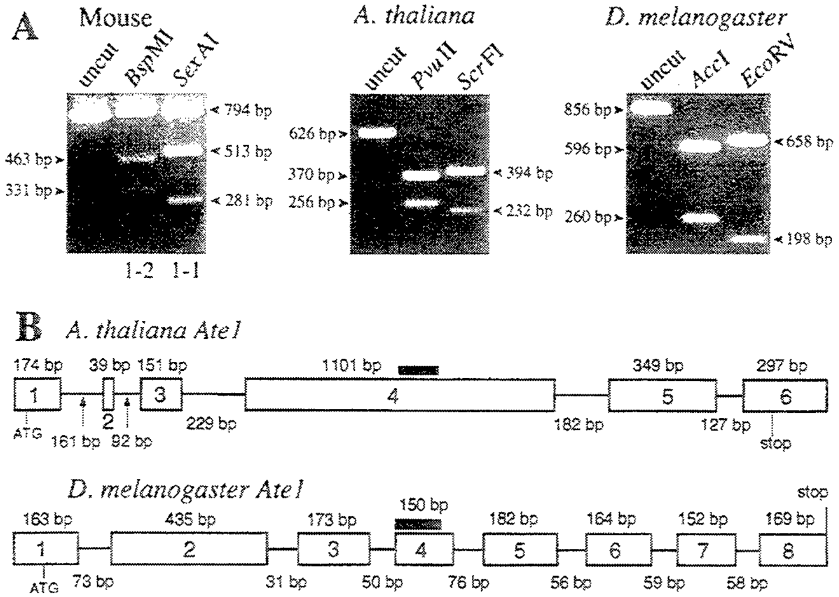


FIG. 5. Alternative splicing of *Ate1* pre-mRNA in mammals (the mouse) but in neither plants (*A. thaliana*) nor arthropods (*D. melanogaster*). (A) The relevant *Ate1* cDNA fragments from mouse (794 bp), *A. thaliana* (626 bp), and *D. melanogaster* (856 bp) were produced by RT-PCR (see Materials and Methods). The products were treated with the indicated restriction endonucleases that cut exclusively within the two alternative 129-bp exons of mouse *Ate1* cDNAs (*Bsp*MI for *Ate1-2*; *Sex*AI for *Ate1-1*) or within the corresponding regions of *A. thaliana* (*Pvu*II and *Scr*FI) and *D. melanogaster* (*Acc*I and *Eco*RV) *Ate1* cDNAs. (B) The *A. thaliana* and *D. melanogaster Ate1* genes. The exon-intron organization of these genes was deduced through comparisons of their cDNA sequences, determined in this work (see Materials and Methods), with the concurrently determined sequences of the corresponding genomic DNA regions (see text). The horizontal lines and rectangles denote, respectively, introns and exons, whose lengths are indicated below and above the line denoting introns. Thick horizontal lines indicate the regions of *A. thaliana* and *D. melanogaster* cDNAs that correspond to the alternative 129-bp exons of the mouse and human *Ate1* cDNAs (Fig. 2 and 4). The lengths of the *A. thaliana* and *D. melanogaster Ate1* genes are, respectively, ~3 and ~2.5 kb.

the *A. thaliana* and *D. melanogaster Ate1* loci, determined through the corresponding sequencing projects, were deposited in GenBank (accession no. AA005237 and accession no. AC004321, respectively). By comparing the cloned *Ate1* cDNAs (see Materials and Methods) and the corresponding genomic sequences of *A. thaliana* and *D. melanogaster*, we could deduce the organization of these *Ate1* genes. The results (Fig. 5B) directly confirmed the absence of the alternative homologous exons from *Ate1* of *A. thaliana* and *D. melanogaster*, in contrast to mammalian *Ate1*. The corresponding region of plant Ate1p is more similar to the exon-encoded sequence of mouse ATE1-1p, whereas in *Drosophila* this region is more similar to the alternative sequence of ATE1-2p (Fig. 2C). The corresponding regions of *S. cerevisiae* and *Schizosaccharomyces pombe* Ate1p are not preferentially similar to either of the two alternative exon-encoding sequences of mouse ATE1p (Fig. 2C).

Mouse ATE1-1p and ATE1-2p can implement the Asp/Glu-specific subset of the N-end rule pathway but differ in activity. To determine whether the two putative mouse R-transferases are in fact R-transferases and to compare their activities in an in vivo setting, we examined whether ATE1-1p and ATE1-2p could confer metabolic instability on Asp-βgal and Glu-βgal in *ate1*Δ *S. cerevisiae*. Asp and Glu are secondary destabilizing residues in the N-end rule (Fig. 1 and introduction). The test substrates Asp-βgal and Glu-βgal (produced through cotranslational deubiquitylation of Ub-Asp-βgal and Ub-Glu-βgal (4)) are short-lived in wild-type yeast (half-lives of ~3 and ~30 min, respectively) but long-lived (half-life of >20 h) in *ate1*Δ *S. cerevisiae* that lacks the *ATE1*-encoded yeast R-transferase (5, 7). Previous work (19, 33) has shown that the steady-state level of an X-βgal protein is a sensitive measure of its metabolic stability.

*S. cerevisiae ate1*Δ cells were cotransformed with a pair of plasmids that expressed one of the two putative mouse R-transferases, ATE1-1p or ATE1-2p, and one of several test substrates (as the corresponding Ub fusions): Asp-βgal, Glu-βgal, Arg-βgal, Cys-βgal, or Met-βgal. Met and Cys are stabilizing residues in the yeast N-end rule; Arg is a primary destabilizing residue; Asp and Glu are secondary destabilizing residues (5, 56). Control tests included either the vector alone or a plasmid expressing *S. cerevisiae* Ate1p. The steady-state levels of X-βgal proteins were determined by measuring the enzymatic activity of βgal in yeast extracts. Using this assay, we found that both forms of mouse ATE1p were able to confer metabolic instability on either Asp-βgal or Glu-βgal in *ate1*Δ *S. cerevisiae* (Fig. 6A). ATE1-1p and ATE1-2p destabilized Glu-βgal much less than Asp-βgal (Fig. 6A), consistent with Glu being a less destabilizing residue in the N-end rule than Asp, presumably because of less efficient arginylation of the N-terminal Glu by R-transferases (54). However, while the apparent destabilizing activity of the mouse ATE1-1p R-transferase was only slightly lower than that of *S. cerevisiae* Ate1p (expressed from the identical vector and promoter), the activity of mouse ATE1-2p was significantly lower than that of ATE1-1p (Fig. 6A).

We also asked whether the two forms of mouse ATE1p could influence each other's activity if they were coexpressed in the same cell (such an influence might be expected, for instance, if the active form of R-transferase were a dimer or if the two forms of R-transferase competed for binding to the same component of a pathway). *S. cerevisiae ate1*Δ cells were cotransformed with two plasmids bearing different selectable markers and expressing different combinations of ATE1-1p and ATE1-2p (1+1, 2+2, or 1+2), and also with a plasmid expressing one of the X-βgal test proteins (X = Met, Arg, Cys, Asp, or Glu). Control cells were cotransformed with the two

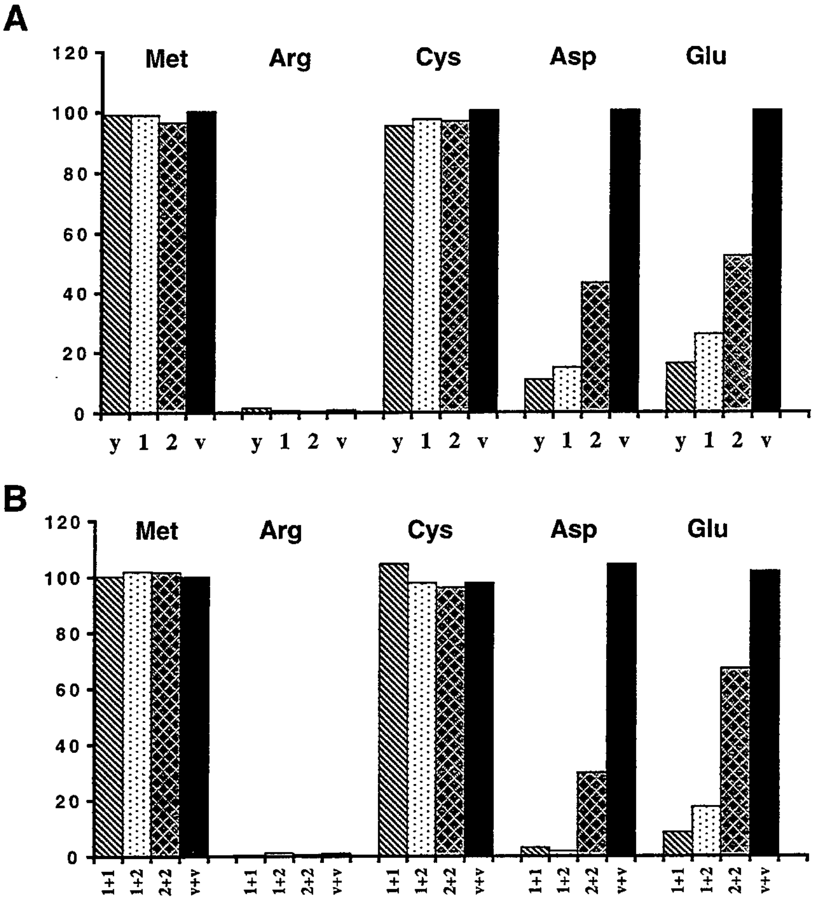


FIG. 6. The two forms of mouse ATE1p can implement the Asp/Glu-specific subset of the N-end rule pathway. (A) Relative enzymatic activities of βgal in *ate1*Δ *S. cerevisiae* transformed with plasmids expressing X-βgal (as Ub-X-βgal) test proteins (X = Met, Arg, Cys, Asp, or Glu) together with a plasmid expressing either yeast *ATE1* (denoted as y), mouse ATE1-1p (denoted as 1), mouse ATE1-2p (denoted as 2), or the vector alone (denoted as v). The N-terminal residues of X-βgals in each set of experiments are indicated at the top. The activity of Met-βgal in cells transformed with vector alone is taken as 100%. (B) The two forms of mouse ATE1p exhibit no cooperativity in mediating the degradation of X-βgals. Shown are relative enzymatic activities of X-βgals in *ate1*Δ *S. cerevisiae* strain cotransformed with plasmids expressing X-βgals (Ub-X-βgals) (X = Met, Arg, Cys, Asp, or Glu) and the combinations of plasmids expressing the following proteins: mouse ATE1-1p and ATE1-1p (1+1), ATE1-1p and ATE1-2p (1+2), ATE1-2p and ATE1-2p (2+2), or two vector controls (v+v). One of the two vectors bore the *TRP1* marker and the other bore the *HIS3* marker (see Materials and Methods). Results are averages of four independent measurements, which differed by less than 10%.

vectors alone. The results (Fig. 6B) indicated that the total activities of the 1+1 and 1+2 combinations (measured as the extent of destabilization of Asp-βgal or Glu-βgal) were similar to each other and much higher than the total activities of 2+2 (Fig. 6B), consistent with the conjecture that the two forms of mouse R-transferase do not interact and that ATE1-1p is a more (possibly much more) active enzyme than ATE1-2p.

Neither ATE1-1p nor ATE1-2p confers metabolic instability on Cys-βgal. Cysteine is a stabilizing residue in the yeast N-end rule but a secondary destabilizing residue in multicellular organisms such as mammals and amphibians (14, 18, 30, 54). The presence of two alternative regions in the two forms of mouse R-transferase (Fig. 4C) initially suggested that one R-transferase might be specific for N-terminal Asp and Glu, with the other specific for Cys. However, Cys-βgal, which is long-lived in wild-type *S. cerevisiae* (5), remained long-lived in the presence of either ATE1-1p or ATE1-2p (Fig. 6A). This finding and, more directly, the results of amino acid sequencing (see below) suggest the existence of a mammalian tRNA-dependent enzyme (presumably a distinct R-transferase) (18) that mediates destabilizing activity of N-terminal Cys.

Mouse ATE1-1p and ATE1-2p destabilize Asp-βgal and Glu-βgal through arginylation of their N-terminal residues. To verify directly that mouse ATE1-1p and ATE1-2p in fact possess the R-transferase activity, we constructed the *ate1*Δ *ubr1*Δ *S. cerevisiae* double mutant AVY34, which lacked both R-transferase and N-recognin (E3), the main recognition component of the N-end rule pathway (see Materials and Methods). Consequently, N-terminal arginylation of a test protein in this mutant by an exogenous R-transferase would not result in degradation of the protein, thereby making it possible to isolate enough of the test protein for N-terminal sequencing. Strain AVY34 was transformed with pUB23-D (expressing Ub-Asp-βgal) and also with either pAT1 (expressing ATE1-1p), pAT2 (expressing ATE1-2p), or vector alone and was grown in SG medium. Asp-βgal proteins isolated from these transformants were subjected to N-terminal sequencing (see Materials and Methods). The results (Table 1) directly confirmed that both ATE1-1p and ATE1-2p possessed R-transferase activity. In agreement with the finding that ATE1-1p was more active than ATE1-2p in destabilizing Asp-βgal in vivo (Fig. 6A), Asp-βgal from cells expressing ATE1-1p was

TABLE 1. N-terminal sequencing of X-βgal proteins isolated from *ate1Δ ubr1Δ S. cerevisiae* expressing different R-transferases

Substrate	Coexpressed protein	N-terminal sequence	Yield (%)
D-e^{K}-βgal		D-H-G-S-A-	
D-e^{K}-βgal	Vector alone	D-H-G-S-A-	~100
D-e^{K}-βgal	Mouse ATE1-1p	R-D-H-G-S-A-	~100
D-e^{K}-βgal	Mouse ATE1-2p	R-D-H-G-S-A-	~50
		D-H-G-S-A-	~50
C-e^{K}-βgal		C-H-G-S-A-	
C-e^{K}-βgal	Vector alone	C-H-G-S-A-	~20
C-e^{K}-βgal	Mouse ATE1-1p	C-H-G-S-A-	~20
C-e^{K}-βgal	Mouse ATE1-2p	C-H-G-S-A-	~20

found to be completely arginylated, whereas Asp-βgal from cells expressing ATE1-2p was arginylated to approximately 50% (Table 1).

We also determined, using the above procedure, whether mouse ATE1-1p or ATE1-2p could arginylate N-terminal Cys. Approximately 80% of Cys-βgal isolated from *ate1Δ ubr1Δ S. cerevisiae* was found to be N-terminally blocked, presumably acetylated (Table 1). However, the rest of Cys-βgal (~20%) bore the N-terminal sequence beginning with Cys and lacking N-terminal Arg, in agreement with the results of the in vivo Cys-βgal degradation assays (Fig. 6 and Table 1). Thus, both ATE1-1p and Ate1-2p are apparently unable to utilize N-terminal Cys as a substrate in *S. cerevisiae*.

ATE1-2p is exclusively cytosolic, whereas ATE1-1p is present in either the nucleus or the cytosol. To determine the intracellular location of the two forms of mouse R-transferase, we constructed fusions to the N terminus of GFP and transiently expressed them in NIH 3T3 cells. Whereas the free 26-kDa GFP was located in both the nucleus and the cytosol (data not shown), the 85-kDa Ate1-2p–GFP fusion was exclusively cytosolic in all of the many transfected cells examined (Fig. 7a to c). In contrast, the 85-kDa ATE1-1p–GFP (the alternative form of R-transferase that is much more active enzymatically than ATE1-2p) was found to be localized differently in different cells on the same coverslip, possibly depending on their cell cycle position and/or metabolic state. Specifically, in ~50% of the transfected cells, ATE1-1p–GFP was exclusively cytosolic (Fig. 7d and e), as was ATE1-2p–GFP (Fig. 7a to c), but in the other ~50% of cells, ATE1-1p–GFP was present in the nucleus as well and, moreover, appeared to be significantly enriched in the nucleus (Fig. 7f and g). Thus, the two 43-residue alternative regions in ATE1-1p and ATE1-2p confer overlapping but nonidentical intracellular distributions on the respective R-transferases.

While the nonuniformity of the ATE1-1p–GFP localization among mouse cells in a single culture remains to be understood, its preferential location in the nuclei of some cells is consistent with a high content of basic residues in its 43-residue region, in comparison to the alternative homologous region of ATE1-2p (Fig. 2C). (No sequences fitting the consensus sequences of known nuclear localization signals could be detected in the 43-residue region of ATE1-1p). In contrast to mouse R-transferases, *S. cerevisiae* Ate1p was shown to be located predominantly in the nuclei of yeast cells (56a).

The ratio of *Ate1-1* to *Ate1-2* mRNA varies greatly among mouse tissues. Northern hybridization, using the 1.1-kb mouse *Ate1* cDNA fragment (nt 638 to 1734) as a probe, detected a

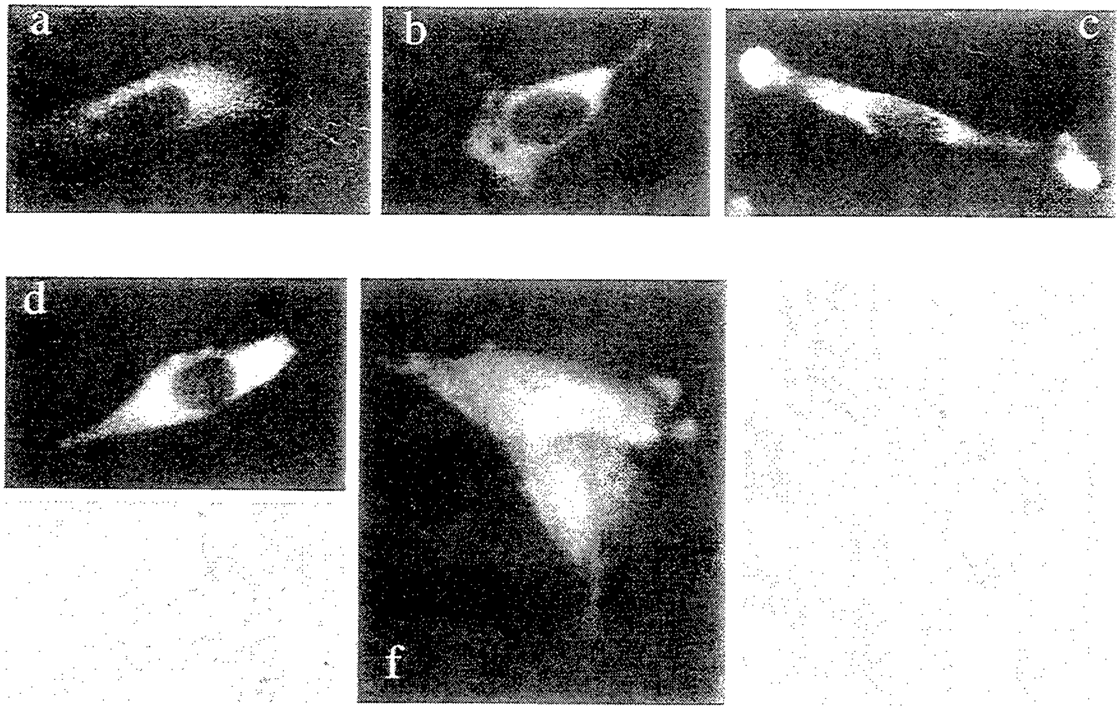


FIG. 7. Intracellular localization of mouse ATE1-1p and ATE1-2p. Shown are green (GFP) fluorescence (a to d and f) and phase-contrast (e and g) micrographs of mouse NIH 3T3 cells transiently transfected with ATE1-1p-GFP (d to g) or ATE1-2p–GFP (a to c) fusion proteins (see Materials and Methods). Panels a to c show different examples of the exclusively cytosolic localization of ATE1-2p. Regions around the nucleus and in the lamellar protrusions at the edges of a cell (c) exhibit higher GFP fluorescence, possibly because of a greater thickness of cells in these areas. Panels d plus e and f plus g show pairs of GFP fluorescence and phase-contrast pictures of cells that express ATE1-1p-GFP. The cell in panels d and e shows ATE1-1p-GFP in the cytosol but not in the nucleus. Cells in panels f and g contain ATE1-1p-GFP in both the cytosol and the nucleus, the latter being apparently enriched in ATE1-lp-GFP.

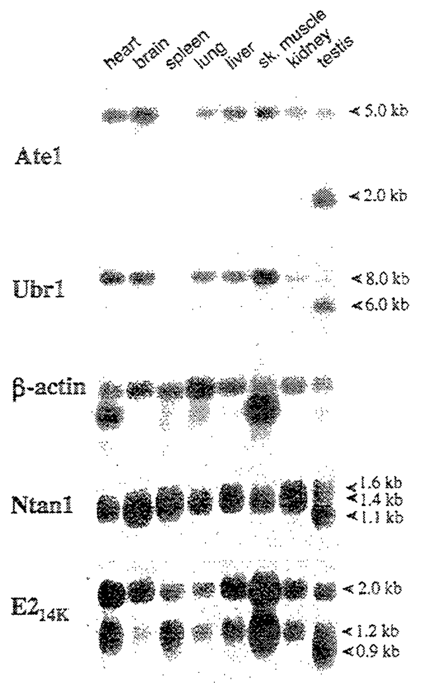


FIG. 8. Northern hybridization analyses of mouse *Ate1* mRNA. The Northern blots of mRNA from different mouse tissues were probed with an *Ate1* cDNA fragment (nt 638 to 1734) which can hybridize to both forms of *Ate1* mRNA. A mouse β-actin cDNA probe was used for comparing the total RNA loads as described previously (19). The same blot was also hybridized with mouse *Ubr1* (29) (see the introduction). The apparent absence of *Ate1* and *Ubr1* mRNAs from the spleen is an artifact of RNA degradation in this lane of the blot (data not shown). Also shown are the results of analogous Northern hybridizations of the mouse *Ntan1* cDNA, encoding the Asn-specific Nt-amidase (Fig. 1A), and E2$_{14K}$, encoding the relevant Ub-conjugating (E2) enzyme (19). The approximate sizes of transcripts are indicated on the right.

single ~5.0-kb transcript (a mixture of the *Ate1-1* and *Ate1-2* mRNAs) in all of the mouse tissues examined except the testis, where the ~5-kb *Ate1* mRNA was a minor one, the major species being ~2 kb (Fig. 8). Both ATE1p and the other targeting components of the mammalian N-end rule pathway are expressed ubiquitously (at various levels), and the testis-specific patterns of transcripts are characteristic for all of them as well (Fig. 8). The existence of the Y-chromosome-encoded, testis-specific variant of the Ub-activating (E1) enzyme (27, 35) suggests that the testis-specific modifications of the N-end rule pathway may be functionally relevant in spermatogenesis.

To determine the ratio of *Ate1-1* to *Ate1-2* mRNA in different mouse tissues or cells in culture, we employed RT-PCR, using sequence differences between the two alternative, homologous 129-bp exons to distinguish between them (see Materials and Methods) (Fig. 3). Approximately equal amounts of *Ate1-1* and *Ate1-2* mRNAs were present in mouse embryonic fibroblasts and in human 293 cells in culture (Fig. 3 and data not shown). However, the molar ratio of *Ate1-1* to *Ate1-2* mRNA was found to vary greatly among the mouse tissues: it was ~0.1 in the skeletal muscle, ~0.25 in the spleen, ~3.3 in the liver, and brain, and ~10 in the testis (Fig. 3). Thus, while the total expression of *Ate1* (*Ate1-1* plus *Ate1-2*) varies by 2- to 4-fold among mouse tissues (Fig. 8), the difference in expression levels between *Ate1-1* and *Ate1-2* mRNAs can be as high as a 100-fold (the skeletal muscle versus the testis) (Fig. 3), suggesting that the two forms of R-transferase may be functionally distinct.

DISCUSSION

The N-end rule pathway is one of several proteolytic pathways of the Ub system (23, 54, 55). Among the targets of the N-end rule pathway are proteins that bear destabilizing N-terminal residues. In the yeast *S. cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by a specific amidase (6), to the secondary destabilizing N-terminal residues Asp and Glu. The destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *ATE1*-encoded R-transferase, to Arg, one of the primary destabilizing residues (7) (Fig. 1B). In mammals, the set of secondary destabilizing residues contains not only Asp and Glu but also Cys, which is a stabilizing residue in yeast (18, 54) (Fig. 1).

In this work, we isolated cDNA encoding the mouse R-transferase, ATE1p, and found that this enzyme exists in two forms, termed ATE1-1p and ATE1-2p, which differ by containing one of the two alternative, homologous 43-residue regions. The two 516-residue R-transferases are produced from the mouse *Ate1* gene by a pathway of alternative splicing that retains one or the other of the two homologous 129-bp exons. The presence of two adjacent, homologous, equal-length, and alternatively utilized exons in a gene (Fig. 4) is nearly unprecedented. To our knowledge, just one such case was described previously: the mouse κE2 enhancer-binding protein E12/E47 (37). The two κE2-binding proteins, E12 and E47, are produced through a switch between two alternative, equal-length exons, resulting in two helix-loop-helix DNA-binding proteins that differ in the ability to homodimerize. Specifically, E47 can bind to the κE2 enhancer either as a homodimer or as a heterodimer with MyoD, whereas E12 can bind as a heterodimer with MyoD but not as a homodimer (37).

We report the following major findings.

(i) Identification, through species walking, and isolation of the mouse cDNA encoding R-transferase (or ATE1p) have shown that mammalian ATE1p exists in two forms, ATE1-1p and ATE1-2p, which differ exclusively by one of the two alternative, homologous 43-residue regions (Fig. 2A).

(ii) The corresponding alternative 129-bp exons are adjacent in the mouse *Ate1* gene. Moreover, the 12-bp sequences around the splice acceptor sites of these exons (6 bp in the intron and 6 bp in the exon) are identical between the two exons (Fig. 4). The splicing of *Ate1* pre-mRNA proceeds in such way that one, and only one, of the alternative 129-bp exons is always retained in the mature *Ate1* mRNA.

(iii) The human *ATE1* gene also contains the two alternative 129-bp exons, whereas the plant (*A. thaliana*) and fly (*D. melanogaster*) *Ate1* genes encode a single form of ATE1p (Fig. 2 and 5). The corresponding 43-residue regions are significantly similar among all of the sequenced R-transferases, from *S. cerevisiae* to mammals (Fig. 2C). The set of *Ate1* genes from mammals to yeast defines a distinct family of proteins, the ATE family. The splicing-derived alternative forms of R-transferase have evolved apparently after the divergence of the arthropod and vertebrate lineages.

(iv) Expression of the mouse *Ate1-1* and *Ate1-2* cDNAs in *ate1*Δ *S. cerevisiae*, and N-terminal sequencing of isolated X-βgal test proteins, was used to show that ATE1-1p and ATE1-2p could implement the Asp/Glu-specific subset of the N-end rule pathway and that they did so through the arginylation of N-terminal Asp or Glu in the test substrates (Fig. 6A and Table 1).

(v) While the destabilizing activity of the mouse ATE1-1p R-transferase is only slightly lower than that of *S. cerevisiae* R-transferase, the activity of mouse ATE1-2p is significantly (possibly considerably) lower than that of ATE1-1p. This conclusion follows also from a comparison of the N-terminal arginylation of Asp-βgal by the two R-transferases (Table 1). The results of coexpressing mouse ATE1-1p and ATE1-2p in the same *ate1*Δ yeast cells were consistent with the conjecture that R-transferase functions as a monomer (Fig. 6B).

(vi) Neither ATE1-1p nor ATE1-2p could confer instability on (or arginylate) Cys-βgal in *ate1*Δ *S. cerevisiae* (Fig. 6A and Table 1). Cys is a stabilizing residue in yeast but a secondary destabilizing residue in the mammalian N-end rule (54). A distinct Cys-specific mammalian R-transferase suggested by these data remains to be identified.

(vii) Mouse ATE1-2p (tested as a GFP fusion) was exclusively cytosolic in mouse 3T3 cells, whereas ATE1-1p was localized differentially in different cells of the same (unsynchronized) culture: it was either exclusively cytosolic or present in both the cytosol and the nucleus (Fig. 7).

(viii) Mouse *Ate1* is a ubiquitously expressed gene. A single ~5-kb mRNA was present in all of the tissues examined except the testis, where the major *Ate1* transcript was ~2 kb in length (Fig. 8). The testis-specific differential expression patterns are also characteristic of the other targeting components of the mammalian N-end rule pathway, such as the *Ntan1*-encoded Asn-specific Nt-amidase and the *Ubr1*-encoded N-recognin (E3α) (19, 29).

(ix) The molar ratio of *Ate1-1* to *Ate1-2* mRNA varies up to a 100-fold among different mouse tissues (Fig. 3 and Results), suggesting a functional significance of the difference between the two R-transferases.

The region of ATE1p that corresponds to the two 129-bp mammalian *Ate1* exons has been significantly conserved throughout eukaryotic evolution, Tyr-296, Gln-297, and His-301 of the mouse ATE1-1p being among the most highly conserved residues (Fig. 2C). No putative members of the ATE family could be detected among the currently known prokaryotic ORFs. The most highly conserved region of R-transferases is an 82-residue stretch (residues 336 to 417) of mouse ATE1p: this region is 95, 76, and 63% identical to the corresponding regions of the human, *D. melanogaster*, and *A. thaliana* ATE1p, respectively (Fig. 2B). A Cys residue(s) is likely to be a component of the active site of R-transferase (31, 32). Among the five fully conserved Cys residues in proteins of the ATE family, four are located in the 56-residue N-terminal region (residues 23 to 78 of mouse ATE1p) (Fig. 2B). Conversion of some of these cysteines in *S. cerevisiae* Ate1p to alanines was found to decrease greatly the R-transferase activity of yeast Ate1p (16a, 32). Furthermore, derivatives of mouse ATE1-1p and ATE1-2p that lacked the first 42 residues were completely inactive in the yeast-based Asp-βgal degradation assay of a kind described in Fig. 7 (data not shown). Finally, a 90-residue C-terminal truncation of *S. cerevisiae* Ate1p did not result in a major decrease of its R-transferase activity (16a). Thus, the active site of R-transferase is likely to encompass at least some of the above N-terminal cysteines.

Since the two mammalian R-transferases (Fig. 4C) are identical in size and, except for a 43-residue region, are identical otherwise as well, it is likely that the previously described (partially purified) mammalian R-transferases (12, 47) were in fact mixtures of Ate1-1p and Ate1-2p. On the other hand, fractionation of a crude R-transferase preparation from rabbit reticulocytes did yield, in addition to a major fraction of R-transferase, chromatographically distinct R-transferase fractions that were not investigated further (12). The ratio of Ate1-1p to Ate1-2p in rabbit reticulocytes is currently unknown.

A splicing-mediated switch that replaces the 129-bp *Ate1* exon of ATE1-1p with the alternative 129-bp exon results in a protein, ATE1-2p, that has a significantly (possibly considerably) lower R-transferase activity (Fig. 6). In addition, ATE1-2p is unable to enter the nucleus (as a GFP fusion), in contrast to ATE1-1p (Fig. 7). Taken together with the finding that the expression ratio of the two *Ate1* mRNAs, *Ate1-1* and *Ate1-2*, varies up to a 100-fold among different mouse tissues (Fig. 3 and Results), these data suggest that the two R-transferases are functionally distinct as well. Cited below are some of the possibilities that are consistent with the available evidence.

Mouse ATE1-2p has the R-transferase activity but arginylates, at steady state, only ~50% of Asp-βgal in *ate1*Δ *S. cerevisiae*, in contrast to both ATE1-2p and *S. cerevisiae* Ate1p (Fig. 6 and Table 1). Moreover, the inefficient arginylation by ATE1-2p occurs in spite of its overexpression in *S. cerevisiae*. In contrast, the yeast Ate1p, which in wild-type *S. cerevisiae* is a weakly expressed protein (7), can quantitatively arginylate in vivo an overexpressed substrate such as Asp-βgal (18). Thus, at a low level of expression (which is likely to be the case in the mouse), the ATE1-2p R-transferase may be, in effect, an inactive enzyme, in contrast to ATE1-1p. If so, ATE1-2p might act as an (indirect) inhibitor of the ATE1-1p function, for example, through a competition with ATE1-1p for the binding to a component of the targeting complex in the N-end rule pathway. (The apparent absence of such competition in *S. cerevisiae* [Fig. 6B] may result from the lack of binding by ATE1p to heterologous yeast proteins.) It is also possible that a large difference in activity between mouse ATE1-1p and ATE1-2p in yeast reflects not their different enzymatic activities in the mouse but a (physiologically irrelevant) differential recognition of an essential yeast cofactor such as Arg-tRNA. Direct comparisons of arginylation kinetics by the purified mouse and yeast R-transferases will be required to address this unlikely but unexcluded interpretation.

Another possibility is that ATE1-2p has a distinct enzymatic activity that has been missed by the current N-terminal arginylation assay (Fig. 6 and Table 1). For example, ATE1-2p might be able to arginylate an internal residue in a substrate protein. In vitro enzymological dissection of ATE1-1p and ATE1-2p will address this and related conjectures. Yet another possibility is that the alternative 43-residue regions of ATE1-1p and ATE1-2p confer different metabolic stabilities on the two R-transferases, the lower apparent activity of ATE1-2p in yeast being due at least in part to its shorter half-life. A test of this model in mouse cells requires antibodies specific for the alternative regions of the two R-transferases; preparation of such antibodies is under way.

In yeast, the N-end rule pathway is present in both the cytosol and the nucleus. The apparent exclusion of mouse ATE1-2p from the nucleus and the different ratios of *Ate1-1* to *Ate1-2* mRNA among the mouse tissues suggest that the rule book of the N-end rule pathway may be regulated differentially in the cytosol and the nucleus, through a cell-type-specific expression of the pathway's components that are located in one but not the other compartment.

Physiological substrates for either eukaryotic R-transferases (54) or their prokaryotic counterparts, L, F-transferases (1, 25, 45) are not known. The cloning and characterization of the first mammalian *Ate1* cDNAs and genes (Fig. 2), and the discovery of alternative splicing that yields mouse ATE1-1p and ATE1-2p (Fig. 4) should facilitate understanding of the functions of mammalian R-transferases, in part through the analysis of ATE1-1p and ATE1-2p enzymes and also because it is now possible to construct mouse strains that lack ATE1-1p and/or ATE1-2p.

ACKNOWLEDGMENTS

The first two authors contributed equally to this work.

We thank Gary Hathaway of the Caltech Microchemistry Facility for the sequencing of X-βgal proteins. We are grateful to Hai Rao, Glenn Turner, Fangyong Du, and Lawrence Peck for helpful suggestions and to Fangyong Du, Federico Navarro-Garcia, Hai Rao, and Youming Xie for comments on the manuscript.

This work was supported by grants DK39520 and GM31530 to A.V. from the National Institutes of Health.

REFERENCES

1. **Abramochkin, G., and T. E. Shrader.** 1995. The leucyl/phenylalanyl-tRNA-protein transferase. Overexpression and characterization of substrate recognition, domain structure, and secondary structure. J. Biol. Chem. **270:**20621–20628.
2. **Alagramam, K., F. Naider, and J. M. Becker.** 1995. A recognition component of the ubiquitin system is required for peptide transport in *Saccharomyces cerevisiae*. Mol. Microbiol. **15:**225–234.
3. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. Struhl (ed.).** 1996. Current protocols in molecular biology. Wiley-Interscience, New York, N.Y.
4. **Bachmair, A., D. Finley, and A. Varshavsky.** 1986. *In vivo* half-life of a protein is a function of its amino-terminal residue. Science **234:**179–186.
5. **Bachmair, A., and A. Varshavsky.** 1989. The degradation signal in a short-lived protein. Cell **56:**1019–1032.
6. **Baker, R. T., and A. Varshavsky.** 1995. Yeast N-terminal amidase: a new enzyme and component of the N-end rule pathway. J. Biol. Chem. **270:**12065–12074.
7. **Balzi, E., M. Choder, W. Chen, A. Varshavsky, and A. Goffeau.** 1990. Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. J. Biol. Chem. **265:**7464–7471.
8. **Bartel, B., I. Wünning, and A. Varshavsky.** 1990. The recognition component of the N-end rule pathway. EMBO J. **9:**3179–3189.
9. **Baumeister, W., J. Walz, F. Zühl, and E. Seemüller.** 1998. The proteasome: paradigm of a self-compartmentalizing protease. Cell **92:**367–380.
10. **Byrd, C., G. C. Turner, and A. Varshavsky.** 1998. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. EMBO J. **17:**269–277.
11. **Chau, V., J. W. Tobias, A. Bachmair, D. Marriott, D. J. Ecker, D. K. Gonda, and A. Varshavsky.** 1989. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science **243:**1576–1583.
12. **Ciechanover, A., S. Ferber, D. Ganoth, S. Elias, A. Hershko, and S. Arfin.** 1988. Purification and characterization of arginyl-tRNA-protein transferase from rabbit reticulocytes. J. Biol. Chem. **263:**11155–11167.
13. **Coux, O., K. Tanaka, and A. L. Goldberg.** 1996. Structure and functions of the 20S and 26S proteasomes. Annu. Rev. Biochem. **65:**801–817.
14. **Davydov, I. V., D. Patra, and A. Varshavsky.** The N-end rule pathway in *Xenopus* oocyte extracts. Arch. Biochem. Biophys., in press.
15. **Dayal, V. K., G. Chakraborty, J. A. Sturman, and N. A. Ingoglia.** 1990. The site of amino acid addition to posttranslationally modified proteins in regenerating rat sciatic nerves. Biochim. Biophys. Acta **1038:**172–177.
16. **deGroot, R. J., T. Rümenapf, R. J. Kuhn, and J. H. Strauss.** 1991. Sindbis virus RNA polymerase is degraded by the N-end rule pathway. Proc. Natl. Acad. Sci. USA **88:**8967–8971.

16a. **Du, F., and A. Varshavsky.** Unpublished data.

17. **Ferber, S., and A. Ciechanover.** 1987. Role of arginine-tRNA in protein degradation by the ubiquitin pathway. Nature **326:**808–811.
18. **Gonda, D. K., A. Bachmair, I. Wünning, J. W. Tobias, W. S. Lane, and A. Varshavsky.** 1989. Universality and structure of the N-end rule. J. Biol. Chem. **264:**16700–16712.
19. **Grigoryev, S., A. E. Stewart, Y. T. Kwon, S. M. Arfin, R. A. Bradshaw, N. A. Jenkins, N. J. Copeland, and A. Varshavsky.** 1996. A mouse amidase specific for N-terminal asparagine: the gene, the enzyme, and their function in the N-end rule pathway. J. Biol. Chem. **271:**28521–28532.

19a. **Grigoryev, S., and A. Varshavsky.** Unpublished data.

20. **Haas, A. J., and T. J. Siepman.** 1997. Pathways of ubiquitin conjugation. FASEB J. **11:**1257–1268.
21. **Hershko, A.** 1991. The ubiquitin pathway for protein degradation. Trends Biochem. Sci. **16:**265–268.
22. **Hill, C. P., N. L. Johnston, and R. E. Cohen.** 1993. Crystal structure of a ubiquitin-dependent degradation substrate: a three-disulfide form of lysozyme. Proc. Natl. Acad. Sci. USA **90:**4136–4140.
23. **Hochstrasser, M.** 1996. Ubiquitin-dependent protein degradation. Annu. Rev. Genet. **30:**405–439.
24. **Hondermarck, H., J. Sy, R. A. Bradshaw, and S. M. Arfin.** 1992. Dipeptide inhibitors of ubiquitin-mediated protein turnover prevent growth factor-induced neurite outgrowth in rat pheochromocytoma PC12 cells. Biochem. Biophys. Res. Commun. **30:**280–288.
25. **Ichetovkin, I. L., G. Abramochkin, and T. E. Shrader.** 1997. Substrate recognition by the leucyl/phenylalanyl-tRNA protein transferase: conservation within the enzyme family and localization to the trypsin-resistant domain. J. Biol. Chem. **272:**33009–33014.

25a. **Johnston, J., and A. Varshavsky.** Unpublished data.

26. **Kaiji, H., G. D. Novelli, and A. Kaiji.** 1963. A soluble amino acid-incorporating system from rat liver. Biochim. Biophys. Acta **76:**474–479.
27. **Kay, G. F., A. Ashworth, G. D. Penny, M. Dunlop, S. Swift, N. Brockdorff, and S. Rastan.** 1991. A candidate spermatogenesis gene on the mouse Y chromosome is homologous to ubiquitin-activating enzyme E1. Nature **354:**486–489.
28. **King, R. W., R. J. Deshaies, J. M. Peters, and M. W. Kirschner.** 1996. How proteolysis drives the cell cycle. Science **274:**1652–1659.

28a. **Kwon, Y. T., V. Denenberg, and A. Varshavsky.** Unpublished data.

29. **Kwon, Y. T., Y. Reiss, V. A. Fried, A. Hershko, J. K. Yoon, D. K. Gonda, P. Sangan, N. G. Copeland, N. A. Jenkins, and A. Varshavsky.** 1998. The mouse and human genes encoding the recognition component of the N-end rule pathway. Proc. Natl. Acad. Sci. USA **95:**7898–7903.

29a. **Kwon, Y. T., and A. Varshavsky.** Unpublished data.

30. **Levy, F., N. Johnsson, T. Rümenapf, and A. Varshavsky.** 1996. Using ubiquitin to follow the metabolic fate of a protein. Proc. Natl. Acad. Sci. USA **93:**4907–4912.
31. **Li, J., and C. Pickart.** 1995. Inactivation of arginyl-tRNA protein transferase by a bifunctional arsenoxide: identification of residues proximal to arsenoxide site. Biochemistry **34:**139–147.
32. **Li, J., and C. M. Pickart.** 1995. Binding of phenylarsenoxide to Arg-tRNA-protein transferase is independent of vicinal thiols. Biochemistry **34:**15829–15837.
33. **Madura, K., R. J. Dohmen, and A. Varshavsky.** 1993. N-recognin/Ubc2 interactions in the N-end rule pathway. J. Biol. Chem. **268:**12046–12054.
34. **Madura, K., and A. Varshavsky.** 1994. Degradation of Gα by the N-end rule pathway. Science **265:**1454–1458.
35. **Mitchell, M. J., D. R. Woods, P. K. Tucker, J. S. Opp, and C. E. Bishop.** 1991. Homology of a candidate spermatogenesis gene from the mouse Y chromosome to the ubiquitin-activating enzyme E1. Nature **354:**483–486.
36. **Mumberg, G., R. Müller, and M. Funk.** 1994. Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. Nucleic Acids Res. **22:**5767–5768.
37. **Murre, C., P. McCaw, and D. Baltimore.** 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell **56:**777–783.
38. **Ota, I. M., and A. Varshavsky.** 1993. A yeast protein similar to bacterial two-component regulators. Science **262:**566–569.
39. **Pickart, C. M.** 1997. Targeting of substrates to the 26S proteasome. FASEB J. **11:**1055–1066.

39a. **Rao, H., and A. Varshavsky.** Unpublished data.

40. **Rechsteiner, M., L. Hoffman, and W. Dubiel.** 1993. The multicatalytic and 26S proteases. J. Biol. Chem. **268:**6065–6068.
41. **Reiss, Y., and A. Hershko.** 1990. Affinity purification of ubiquitin-protein ligase on immobilized protein substrates. J. Biol. Chem. **265:**3685–3690.
42. **Rothstein, R.** 1991. Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol. **194:**281–301.
43. **Rubin, D. M., S. van Nocker, M. Glickman, O. Coux, I. Wefes, S. Sadis, H. Fu, A. Goldberg, R. Vierstra, and D. Finley.** 1997. ATPase and ubiquitin-binding proteins of the yeast proteasome. Mol. Biol. Rep. **24:**17–26.
44. **Scheffner, M., S. Smith, and S. Jentsch.** 1998. The ubiquitin conjugation system, p. 65–98. *In* J.-M. Peters, J. R. Harris, and D. Finley (ed.), Ubiquitin and the biology of the cell. Plenum Press, New York, N.Y.
45. **Shrader, T. E., J. W. Tobias, and A. Varshavsky.** 1993. The N-end rule in *Escherichia coli*: cloning and analysis of the leucyl, phenylalanyl-tRNA-protein transferase gene *aat*. J. Bacteriol. **175:**4364–4374.
46. **Sijst, A. J. A. M., I. Pilip, and E. G. Pamer.** 1997. The *Listeria monocytogenes*-secreted p60 protein is an N-end rule substrate in the cytosol of infected cells. J. Biol. Chem. **272:**19261–19268.
47. **Soffer, R. L.** 1980. Biochemistry and biology of aminoacyl-tRNA-protein transferases, p. 493–505. *In* D. Söll, J. Abelson, and P. R. Schimmel (ed.), Transfer RNA: biological aspects. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
48. **Solomon, V., V. Baracos, P. Sarraf, and A. Goldberg.** When muscles atrophy, rates of ubiquitin conjugation increase, largely through activation of the N-end rule pathway. Proc. Natl. Acad. Sci. USA, in press.
49. **Solomon, V., S. H. Lecker, and A. L. Goldberg.** 1998. The N-end rule pathway mediates a major fraction of protein degradation in skeletal muscle. J. Biol. Chem. **273:**25216–2522.
50. **Stewart, A.** 1995. Trends in genetics nomenclature guide. Elsevier Science, Ltd., Cambridge, United Kingdom.
51. **Stewart, A. E., S. M. Arfin, and R. A. Bradshaw.** 1995. The sequence of porcine protein N-terminal asparagine amidohydrolase: a new component of the N-end rule pathway. J. Biol. Chem. **270:**25–28.
52. **Taban, C. H., H. Hondermarck, R. A. Bradshaw, and B. Boilly.** 1996. Effect of a dipeptide inhibiting ubiquitin-mediated protein degradation on nerve-dependent limb regeneration in the newt. Experientia **52:**865–870.
53. **Tobias, J. W., T. E. Shrader, G. Rocap, and A. Varshavsky.** 1991. The N-end rule in bacteria. Science **254:**1374–1377.

54. **Varshavsky, A.** 1997. The N-end rule pathway of protein degradation. Genes Cells 2:13–28.
55. **Varshavsky, A.** 1997. The ubiquitin system. Trends Biochem. Sci. 22:383–387.
56. **Varshavsky, A., C. Byrd, I. V. Davydov, R. J. Dohmen, F. Du, M. Ghislain, M. Gonzalez, S. Grigoryev, E. S. Johnson, N. Johnsson, J. A. Johnston, Y. T. Kwon, F. Lévy, O. Lomovskaya, K. Madura, I. Ota, T. Rümenapf, T. E. Shrader, T. Suzuki, G. Turner, P. R. H. Waller, and A. Webster.** 1998. The N-end rule pathway, p. 223–278. *In* J.-M. Peters, J. R. Harris, and D. Finley (ed.), Ubiquitin and the biology of the cell. Plenum Press, New York, N.Y.

56a. **Wang, H. R., and A. Varshavsky.** Unpublished data.

57. **Wang, Y. M., and N. A. Ingoglia.** 1997. N-terminal arginylation of sciatic nerve and brain proteins following injury. Neurochem. Res. 22:1453–1459.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1999 by The American Society for Biochemistry and Molecular Biology, Inc.
Vol. 274, No. 25, Issue of June 18, pp. 18135–18139, 1999
Printed in U.S.A.

Bivalent Inhibitor of the N-end Rule Pathway*

(Received for publication, February 9, 1999)

Yong Tae Kwon‡, Frédéric Lévy§, and Alexander Varshavsky‡¶

From the ‡Division of Biology, California Institute of Technology, Pasadena, California 91125
and the §Ludwig Institute for Cancer Research, 155, ch. Des Boveresses, CH-1066 Epalinges, Switzerland

The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Ubr1p, the recognition (E3) component of the *Saccharomyces cerevisiae* N-end rule pathway, contains at least two substrate-binding sites. The type 1 site is specific for N-terminal basic residues Arg, Lys, and His. The type 2 site is specific for N-terminal bulky hydrophobic residues Phe, Leu, Trp, Tyr, and Ile. Previous work has shown that dipeptides bearing either type 1 or type 2 N-terminal residues act as weak but specific inhibitors of the N-end rule pathway. We took advantage of the two-site architecture of Ubr1p to explore the feasibility of bivalent N-end rule inhibitors, whose expected higher efficacy would result from higher affinity of the cooperative (bivalent) binding to Ubr1p. The inhibitor comprised mixed tetramers of β-galactosidase that bore *both* N-terminal Arg (type 1 residue) and N-terminal Leu (type 2 residue) but that were resistant to proteolysis *in vivo*. Expression of these constructs in *S. cerevisiae* inhibited the N-end rule pathway much more strongly than the expression of otherwise identical β-galactosidase tetramers whose N-terminal residues were exclusively Arg or exclusively Leu. In addition to demonstrating spatial proximity between the type 1 and type 2 substrate-binding sites of Ubr1p, these results provide a route to high affinity inhibitors of the N-end rule pathway.

Among the targets of the N-end rule pathway are intracellular proteins bearing destabilizing N-terminal residues (1, 2). This proteolytic pathway is one of several pathways of the ubiquitin (Ub)[1] system, whose diverse functions include the regulation of cell growth, division, differentiation, and responses to stress (3–6). Ub is a 76-residue eukaryotic protein that exists in cells either free or conjugated to other proteins. Many of the Ub-dependent regulatory circuits involve processive degradation of ubiquitylated proteins by the 26 S proteasome, an ATP-dependent multisubunit protease (7, 8).

The N-end rule is organized hierarchically. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by the *NTA1*-encoded N-terminal amidase, into the secondary destabilizing N-terminal residues Asp and Glu. The destabilizing activity of N-terminal Asp and Glu requires their conjugation by the *ATE1*-encoded Arg-tRNA-protein transferase (R-transferase) to Arg, one of the primary destabilizing residues (reviewed in Refs. 1 and 9). In mammals, two distinct N-terminal amidases specific, respectively, for N-terminal Asn or Gln mediate the conversion of these tertiary destabilizing residues into the secondary destabilizing residues Asp or Glu (10, 11). The set of secondary destabilizing residues in vertebrates contains not only Asp and Glu but also Cys, which is a stabilizing residue in yeast (9, 12, 13).

The primary destabilizing N-terminal residues are bound directly by N-recognin, the E3 (recognition) component of the N-end rule pathway. In *S. cerevisiae*, N-recognin is the *UBR1*-encoded 225-kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues: Phe, Leu, Trp, Tyr, Ile, Arg, Lys, and His (1, 14). The *Ubr1* genes encoding mouse and human N-recognin (also called E3α) have been cloned as well (15). N-recognin has at least two substrate-binding sites. The type 1 site is specific for the basic N-terminal residues Arg, Lys, and His. The type 2 site is specific for the bulky hydrophobic N-terminal residues Phe, Leu, Trp, Tyr, and Ile (1, 12, 16, 17). N-recognin can also target short-lived proteins such as Cup9p (18) and Gpa1p (19, 20), which lack destabilizing N-terminal residues. The Ubr1p-recognized degradation signals of these proteins remain to be characterized in detail.

The known functions of the N-end rule pathway include the control of di- and tripeptide import in *S. cerevisiae* through the degradation of Cup9p, a transcriptional repressor of the peptide transporter gene *PTR2* (18, 21); a mechanistically undefined role in the Sln1p-dependent phosphorylation cascade that mediates osmoregulation in *S. cerevisiae* (22); the degradation of Gpa1p, a Gα protein of *S. cerevisiae* (19, 20); and the conditional degradation of alphaviral RNA polymerase in virus-infected metazoan cells (23). Physiological N-end rule substrates were also identified among the proteins secreted into the cytosol of the host cell by intracellular parasites such as the bacterium *Listeria monocytogenes* (24). Short half-lives of these proteins are required for the efficient presentation of their peptides to the immune system (24). A partial inhibition of the N-end rule pathway was reported to interfere with mammalian cell differentiation (25) and to delay limb regeneration in amphibians (26). Recent evidence suggests that the N-end rule pathway mediates a large fraction of the muscle protein turnover (27) and plays a role in catabolic states that result in muscle atrophy (28).

Targeted mutagenesis has been used to inactivate the N-end rule pathway in *Escherichia coli* and *S. cerevisiae* (14, 29). Analogous mutants have recently been constructed in the mouse as well.[2] These approaches notwithstanding, an efficacious inhibitor of the N-end rule pathway would be useful as well, especially with organisms less tractable genetically. The emerging understanding of the N-end rule pathway in mammals suggests that selective inhibition or activation of this proteolytic system may also have medical applications. Previ-

---

* This work was supported by National Institutes of Health Grant DK39520 (to A. V.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *"advertisement"* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: Div. of Biology, 147-75, Caltech, 1200 East California Blvd., Pasadena, CA 91125. Tel.: 626-395-3785; Fax: 626-440-9821; E-mail: avarsh@its.caltech.edu.

[1] The abbreviations used are: Ub, ubiquitin; βgal, *E. coli* β-galactosidase; E3, ubiquitin-protein ligase; ha, hemagglutinin.

[2] Y. T. Kwon and A. Varshavsky, unpublished data.

ous work has shown that millimolar concentrations of amino acid derivatives such as dipeptides bearing destabilizing N-terminal residues can selectively inhibit the N-end rule pathway in extracts from rabbit reticulocytes (12, 17) and *Xenopus* eggs (13), and in intact *S. cerevisiae* cells as well (16). However, the same dipeptides were observed to have at most marginal effects on the N-end rule pathway in intact mammalian cells.[3] One limitation of dipeptide inhibitors is their apparently low affinity for the type 1 and the type 2 site of N-recognin (30).

In the present work, we explored the possibility that a bivalent ligand can bind simultaneously to the type 1 and type 2 sites of N-recognin (see Fig. 1*A*). Similarly to the previously characterized bivalent interactions that involve either macromolecules or small molecules (31, 32), the cooperativity of binding at two independent, mutually nonexclusive sites would be expected to increase the affinity between N-recognin and a bivalent inhibitor by orders of magnitude, in comparison with the affinity of a monovalent binding by the same compound. We show that a bivalent inhibitor of the N-end rule pathway is feasible and consider the implications of this advance.

EXPERIMENTAL PROCEDURES

Strains and General Techniques—The *S. cerevisiae* strains used were JD52 (*MAT***a** *ura3-52 his3-Δ200 leu2-3,112 trp1-Δ63 lys2-801*) and JD55 (*MAT***a** *ura3-52 his3-Δ200 leu2-3,112 trp1-Δ63 lys2-801 ubr1 Δ::HIS3*) (19, 33). Cells were grown on rich (YPD) or synthetic medium containing either 2% dextrose (SD medium), 2% galactose (SG medium), or 2% raffinose (SR medium) (34). To induce the P_{CUP1} promoter, $CuSO_4$ was added to a final concentration of 0.1 mM. Transformation of *S. cerevisiae* was carried out using the lithium acetate method (35).

Plasmids—The high copy (2μ-based) plasmids pRΔ-βgal-TRP1 and pRΔ-βgal-HIS3, which expressed Arg-$e^{\Delta K}$-βgal (Ub-Arg-$e^{\Delta K}$-βgal) (see Fig. 2*A*) from the galactose-inducible $P_{CYC1/GAL1}$ hybrid promoter (2), were produced by replacing the *URA3* marker gene of pFL7 with either *TRP1* or *HIS3*. pLΔ-βgal-TRP1 and pLΔ-βgal-HIS3, both of which expressed Leu-$e^{\Delta K}$-βgal (Ub-Leu-$e^{\Delta K}$-βgal), were produced by replacing the Ub-Arg domain of pRΔ-βgal-TRP1 and pRΔ-βgal-HIS3 with Ub-Leu domain of the pLL2 plasmid.[4] The plasmid pFL7 was produced from pUB23-R (2) by converting the lysine codons 15 and 17 of the extension e^K into arginine codons (36, 37), yielding a construct encoding the extension $e^{\Delta K}$ in front of a βgal moiety lacking the first 23 residues of wild type βgal (see Fig. 2*A*). The low copy, pRS315 vector-derived (38) plasmid pR-eΔKhaUra3-R3R7 expressed Arg-$e^{\Delta K}$-ha-Ura3p^{K3R,K7R} (Ub-Arg-$e^{\Delta K}$-ha-Ura3p^{K3R,K7R}) from the P_{CUP1} promoter. Arg-$e^{\Delta K}$-ha-Ura3p^{K3R,K7R} (see Fig. 2*B*) is called Arg-Ura3p in the main text. In this N-end rule substrate, the residues Lys-3 and Lys-7 of the *S. cerevisiae* Ura3p were converted to arginines (see "Results and Discussion"). In addition, the ha epitope tag (39) was placed between $e^{\Delta K}$ and Ura3p^{K3R,K7R} (see Fig. 2*B*). The plasmid pR-eΔKhaUra3-R3R7 was produced from pFL1 (encoding Ub-Arg-$e^{\Delta K}$-ha-Ura3p) through site-directed mutagenesis of the *URA3* codons for Lys-3 and Lys-7. pFL1 was produced from pKM1235 (which encoded Ub-Arg-e^K-ha-Ura3p)[5] by converting the e^K-coding sequence into the one encoding $e^{\Delta K}$.

Pulse-Chase and Plating Efficiency Assays—Pulse-chase assays with *S. cerevisiae* in mid-exponential growth (A_{600} of ~1) utilized ^{35}S-EXPRESS (NEN Life Science Products) and were carried out as described previously (10, 19), including the immunoprecipitation with anti-βgal and anti-ha antibodies and quantitation with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA). To determine plating efficiency, *S. cerevisiae* strains JD52 (*UBR1*) and JD55 (*ubr1Δ*) expressing Arg-Ura3p (Ub-Arg-$e^{\Delta K}$-ha-Ura3p^{K3R,K7R}; see Fig. 2*B*) were co-transformed with plasmids indicated in the legend to Fig. 3. The transformants were cultured in the raffinose-based medium (SR) lacking Leu, His, and Trp for 20 h. The cultures were then diluted into the otherwise identical galactose-containing (SG) medium to a final A_{600} of 0.1. At an A_{600} of 0.4, cultures were either supplemented with 0.1 mM $CuSO_4$ or left unsupplemented. At the A_{600} of 1.0, the cultures were diluted with SG (which lacks Leu, His, and Trp) either containing or lacking 0.1 mM $CuSO_4$ and were plated on the plates of the same medium composition that also either contained or lacked uracil. The plating efficiency (%) was defined as the ratio of the number of colonies on SG (−Leu, −His, −Trp, −Ura) plates to the number of colonies on SG (−Leu, −His, −Trp) plates, at the same concentration of $CuSO_4$. For each measurement, colonies on 15 plates were counted to yield the average number of colonies per plate.

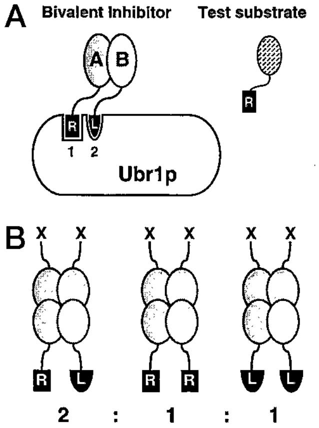


FIG. 1. **The concept of a bivalent inhibitor of the N-end rule pathway.** *A*, the type 1 and type 2 sites of *S. cerevisiae* Ubr1p (N-recognin), which are specific, respectively, for the basic (Arg, Lys, and His) and bulky hydrophobic (Phe, Leu, Trp, Tyr, and Ile) N-terminal residues. In the diagram, the type 1 and type 2 sites are occupied by their ligands, the N-terminal Arg and Leu, borne by a heterodimeric bivalent inhibitor (actually, a tetrameric βgal-based protein in the present work). A test substrate bearing Arg, a type 1 destabilizing N-terminal residue is shown as well. The test substrate, in contrast to the protein-based inhibitor, bears at least one internal Lys residue (not indicated in the diagram) that can function as a component of the N-degron. The type 1 and type 2 sites of N-recognin are shown located close together in the N-terminal region of the 225-kDa Ubr1p. The recent genetic dissection of the Ubr1p substrate-binding sites[6] placed the type 1 and type 2 sites close together in the ~60-kDa N-terminal region of the 225-kDa Ubr1p. *B*, a diagram illustrating the expected frequencies of heterodimeric (Arg- and Leu-bearing) dimers within a βgal-based bivalent inhibitor. Specifically, at equal levels of expression of the two βgal-based polypeptide chains, 50% of βgal tetramers would be expected to be heterotetramers in which at least one of the two dimers bears different (Arg and Leu) N-terminal residues. In the βgal tetramer, the two N termini of each dimer are spatially close, exposed, and oriented in the same direction (40). See also "Results and Discussion."

RESULTS AND DISCUSSION

We constructed a bivalent N-end rule inhibitor (Fig. 1*A*) from the previously studied N-end rule substrates derived from *E. coli* βgal (2). In eukaryotes, linear Ub-protein fusions are rapidly cleaved by deubiquitylating enzymes at the Ub-protein junction, making possible the production of otherwise identical proteins bearing different N-terminal residues, a technical advance that led to the finding of the N-end rule (2). A βgal-based N-end rule substrate contains a destabilizing N-terminal residue (produced *in vivo* using the Ub fusion technique (1)); a ~45-residue, *E. coli* Lac repressor-derived N-terminal extension called e^K (extension *e* bearing lysines *K*); and the βgal moiety lacking its first 21 residues. The resulting X-e^K-βgal is a short-lived protein in both yeast and mammalian cells, whereas an otherwise identical protein bearing a stabilizing N-terminal residue such as Met or Val is metabolically stable (1, 2). An N-degron comprises a destabilizing N-terminal residue and a Lys residue (or residues), the latter being the site of formation of a multi-Ub chain (1, 36). (Ubr1p can also recognize a set of other, internal degrons, which remain to be characterized (18).) If Lys-15 and Lys-17 of the e^K extension are replaced by the Arg residues (which cannot be ubiquitylated), the resulting X-$e^{\Delta K}$-βgal (Fig. 2*A*) is long-lived *in vivo* even if its N-terminal residue is destabilizing in the N-end rule (1, 37).

In the present work, we used the metabolically stable Arg-

---

[3] F. Lévy and A. Varshavsky, unpublished data.

[4] M. Ghislain and A. Varshavsky, unpublished data.

[5] K. Madura and A. Varshavsky, unpublished data.

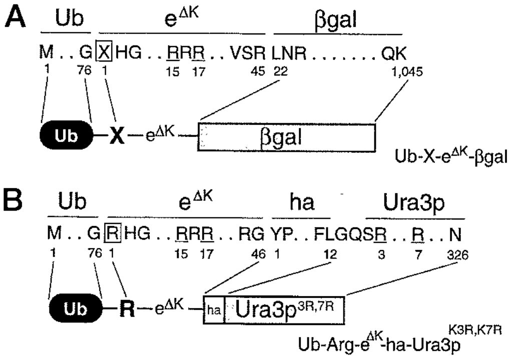

A
Ub
e^ΔK
βgal
M . . G X HG . . RRR . . VSR LNR . . . . QK
1 76 1 15 17 45 22 1,045
Ub
X
e^ΔK
βgal
Ub-X-e^ΔK-βgal
B
Ub
e^ΔK
ha
Ura3p
M . . G R HG . . RRR . . RG YP. . FLGQSR . . R . . N
1 76 1 15 17 46 1 12 3 7 326
Ub
R
e^ΔK
ha
Ura3p^3R,7R
Ub-Arg-e^ΔK-ha-Ura3p^K3R,K7R

FIG. 2. **Designs of bivalent inhibitor and test substrate.** *A*, the βgal-based fusions (the residue *X* was either Arg or Leu) used to construct the Arg/Leu-bearing bivalent inhibitor. The Ub moiety of the fusions was cotranslationally removed *in vivo* by deubiquitinating enzymes (1). The ~45-residue, *E. coli* Lac repressor-derived sequence termed $e^{\Delta K}$ (extension (*e*) lacking lysines (Δ*K*)), is described in the main text. The βgal part of the fusion lacked the first 21 residues of wild type βgal (2). *B*, the Ura3p-based N-end rule substrate, Arg-$e^{\Delta K}$-ha-Ura3p^{K3R,K7R}, derived from Ub-Arg-$e^{\Delta K}$-ha-Ura3p^{K3R,K7R} and denoted Arg-Ura3p, is described in the main text.

$e^{\Delta K}$-βgal (produced from Ub-Arg-$e^{\Delta K}$-βgal) and Leu-$e^{\Delta K}$-βgal (produced from Ub-Leu-$e^{\Delta K}$-βgal). These proteins retain the ability to bind, respectively, to the type 1 and type 2 sites of N-recognin but cannot be ubiquitylated (37), apparently because the most N-terminal Lys residue in X-$e^{\Delta K}$-βgal, at position 239, is too far from the N terminus of the protein. In the βgal tetramer, the two N termini of each dimer are spatially close, exposed, and oriented in the same direction (40). At equal levels of expression of the two βgal-based polypeptide chains such as Arg-$e^{\Delta K}$-βgal and Leu-$e^{\Delta K}$-βgal, 50% of βgal tetramers would be expected to be heterotetramers in which at least one of the two dimers bears different (Arg and Leu) N-terminal residues (Fig. 1*B*). If the type 1 and type 2 substrate-binding sites of the 225-kDa Ubr1p are appropriately located and oriented, they might be able to bind the Arg- and Leu-bearing subunits of the mixed βgal tetramer, especially in view of the presumed flexibility of the $e^{\Delta K}$ extension (1) (Fig. 1*A*).

The reporter N-end rule substrate in this study was Arg-$e^{\Delta K}$-ha-Ura3p^{K3R,K7R}, denoted below as Arg-Ura3p (Fig. 2*B*). This ha-tagged, type 1 N-end rule substrate was produced from Ub-Arg-$e^{\Delta K}$-ha-Ura3p^{K3R,K7R} through the cotranslational *in vivo* cleavage by deubiquitinating enzymes (1, 6, 41). The lysine-lacking $e^{\Delta K}$ extension of Arg-$e^{\Delta K}$-ha-Ura3p^{K3R,K7R}, and the replacement of the first two lysines of the Ura3p moiety with arginines were used to decrease the rate of degradation of Arg-Ura3p by the N-end rule pathway and also to reduce the slow but detectable degradation of Arg-Ura3p by yet another pathway, through a degron distinct from the N-degron.[3] Several Lys residues of Ura3p other than Lys-3 and Lys-7 are also close to its N terminus, thus accounting for the absence, in this case, of the all-or-none effect on the reporter degradation that is observed when e^K is replaced with $e^{\Delta K}$ in an X-e^K-βgal substrate (37). The Lys-3 → Arg and Lys-7 → Arg modifications decreased the enzymatic activity of the Ura3p moiety.[2] The reduced enzymatic activity of Ura3p^{K3R,K7R} facilitated selection assays (Figs. 3 and 4).

The first bivalent inhibitor assay employed *ura3 S. cerevisiae* expressing Arg-Ura3p (Fig. 2*B*) from the uninduced P_{CUP1} promoter. The Ubr1p-mediated degradation of Arg-Ura3p ($t_{1/2}$ of ~8 min) and its correspondingly low steady-state concentration rendered wild type (*UBR1*) cells phenotypically Ura$^-$, whereas *ubr1*Δ strains expressing Arg-Ura3p were phenotypically Ura$^+$ (Figs. 3 and 4 and data not shown). Cells expressing Arg-Ura3p were cotransformed with two control plasmids (vectors; *423+424* in Fig. 3). Alternatively, these cells were cotransformed with two plasmids (bearing different selectable markers) that expressed either Arg-$e^{\Delta K}$-βgal alone (*R+R* in Fig. 3), Leu-$e^{\Delta K}$-βgal alone (*L+L* in Fig. 3), or both of them together (*R+L* in Fig. 3; the bivalent inhibitor mode) from a galactose-inducible promoter. Pairs of alternatively marked plasmids were used to make certain that the conditions of expression and the total amounts of βgal-based proteins produced remained the same in all of these settings. The transformants were streaked on SG medium lacking uracil.

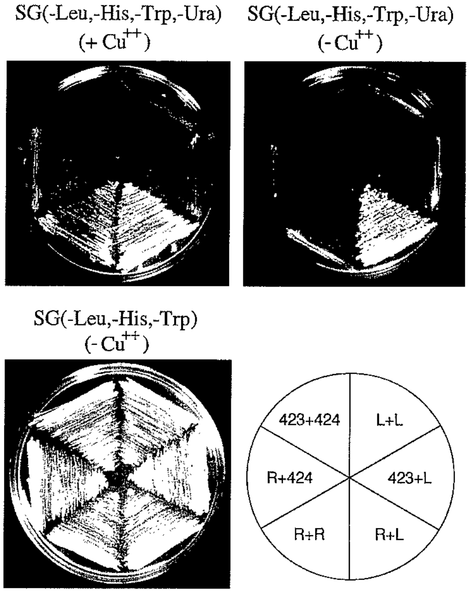


FIG. 3. **A bivalent, but not monovalent, inhibitor of Ubr1p confers Ura$^+$ phenotype on cells expressing the short-lived N-end rule substrate Arg-Ura3p.** *S. cerevisiae* JD52 (*UBR1*) expressing Arg-Ura3p (Ub-Arg-Ura3p), a short-lived reporter (Fig. 2B), were cotransformed, alternatively, with pRS423 (*HIS3*-based control vector) and pRS424 (P_{GAL1}, *TRP1*-based control vector) (denoted as *423+424*); with pRS424 and pRΔ-βgal-HIS3, expressing Arg-$e^{\Delta K}$-βgal (Ub-Arg-$e^{\Delta K}$-βgal) (denoted as *R+424*; monovalent inhibitor); with pRS423 and pLΔ-βgal-TRP1, expressing Leu-$e^{\Delta K}$-βgal (Ub-Leu-$e^{\Delta K}$-βgal) (denoted as *423+L*); with pRΔ-βgal-HIS3 and pRΔ-βgal-TRP1, both expressing Arg-$e^{\Delta K}$-βgal (denoted as *R+R*; monovalent inhibitor); with pLΔ-βgal-HIS3 and pLΔ-βgal-TRP1, both expressing Leu-$e^{\Delta K}$-βgal (denoted as *L+L*); or with pRΔ-βgal-HIS3 and pLΔ-βgal-TRP1, expressing Arg-$e^{\Delta K}$-βgal and Leu-$e^{\Delta K}$-βgal (denoted as *R+L*; the bivalent inhibitor). Cells were streaked on SG medium containing 0.1 mM $CuSO_4$ and lacking Leu, His, Trp, and Ura (*upper left panel*), on the otherwise identical medium lacking the added $CuSO_4$ (*upper right panel*), or on the Ura-containing SG medium lacking Leu, His, and Trp (controls; *lower left panel*). Plates were incubated at 30 °C for 3 days.

Remarkably, only those Arg-Ura3p-expressing cells that expressed *both* Arg-$e^{\Delta K}$-βgal and Leu-$e^{\Delta K}$-βgal became Ura$^+$ under these conditions (Fig. 3). The cells that expressed either Arg-$e^{\Delta K}$-βgal alone or Leu-$e^{\Delta K}$-βgal alone remained Ura$^-$, as did the cells that received control plasmids (Fig. 3). (The same cells grew equally well in the control SG medium containing uracil (Fig. 3, *bottom left panel*).) Note that the monovalent inhibitors were ineffective despite the fact that the concentration of either the Arg-based N terminus alone or the Leu-based N terminus alone was twice the concentration of the same N termini in the case of the bivalent inhibitor.

To quantify the effect of coexpressing Arg-$e^{\Delta K}$-βgal and Leu-$e^{\Delta K}$-βgal on the rescue of the Ura$^+$ phenotype, a plating efficiency assay was carried out with the same transformants. Equal amounts of cells were plated on SG(+Ura) and SG(−Ura) plates, and the numbers of colonies were determined. When the Arg-Ura3p reporter was expressed at a sufficiently low rate (uninduced P_{CUP1} promoter), cells became

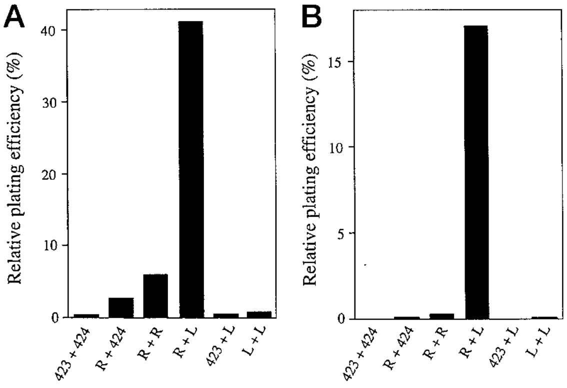


FIG. 4. **Plating efficiencies of *S. cerevisiae* expressing Arg-Ura3p in the presence of bivalent and monovalent inhibitors of the N-end rule pathway.** *A*, *S. cerevisiae* JD52 (*UBR1*) and JD55 (*ubr1Δ*) expressing Arg-Ura3p were cotransformed with the sets of plasmids described and denoted in the legend to Fig. 3. The transformants were cultured as described under "Experimental Procedures" and plated on either SG(−Leu, −His, −Trp, −Ura) plates or control plates SG(−Leu, −His, −Trp) containing 0.1 mM $CuSO_4$. The plating efficiencies shown are the values produced by normalization against the absolute plating efficiency (92%) of the positive control: the *ubr1Δ* strain JD55 expressing Arg-Ura3p and bearing the vector pRS424. *B*, the same experiment was done using plates lacking the added $CuSO_4$. The plating efficiencies shown are the values produced by normalization against the positive control used in *A*. Under these growth conditions (no added $CuSO_4$), the absolute plating efficiency of the positive control was 26%.

Ura^{+} (through metabolic stabilization of Arg-Ura3p) only in the presence of *both* Arg-e$^{\Delta K}$-βgal and Leu-e$^{\Delta K}$-βgal (Fig. 4*B*). A weak stabilizing effect of Arg-e$^{\Delta K}$-βgal alone could be detected only at a ~20-fold higher level of Arg-Ura3p expression (induced P$_{CUP1}$ promoter) (Fig. 4*A*). No stabilization of Arg-Ura3p was observed in the presence of Leu-e$^{\Delta K}$-βgal under any conditions (Fig. 4), confirming the specificity of inhibition in regard to the type (basic or bulky hydrophobic) of the primary destabilizing N-terminal residue of the reporter. Higher sensitivity of this assay at the higher level of Arg-Ura3p expression results from a higher steady-state level of the short-lived Arg-Ura3p, so that even its marginal stabilization suffices to render a small fraction of cells Ura^{+} (Fig. 4*A*; compare with Fig. 4*B*).

To analyze directly the *in vivo* degradation of Arg-Ura3p in the presence of different combinations of X-e$^{\Delta K}$-βgal proteins, the transformants of Figs. 3 and 4 were subjected to pulse-chase analysis, with immunoprecipitation of both Arg-Ura3p and the (long-lived) X-e$^{\Delta K}$-βgals (Fig. 5). Quantitation of the resulting electrophoretic patterns (Fig. 5*C*) confirmed and extended the conclusions reached through phenotypic analyses (Figs. 3 and 4). Specifically, the normally short-lived Arg-Ura3p (Fig. 5*A*, *lanes 1–3*) was strongly (but still incompletely) stabilized in the presence of both Arg-e$^{\Delta K}$-βgal and Leu-e$^{\Delta K}$-βgal (Fig. 5*A*, *lanes 4–6*; compare with *lanes 1–3* and *7–9*). This stabilization was manifested especially clearly as an increase in the relative amount of Arg-Ura3p at the beginning of chase (time 0), indicating reduced degradation of Arg-Ura3p during the pulse (Fig. 5*C*). This latter degradation pattern, termed "zero point effect," is caused by the previously demonstrated preferential targeting of newly formed (as distinguished from conformationally mature) protein substrates by the N-end rule pathway (16, 42). The increased steady-state level of Arg-Ura3p in the presence of both Arg-e$^{\Delta K}$-βgal and Leu-e$^{\Delta K}$-βgal accounted for the results of phenotypic analyses (Figs. 3 and 4). The much smaller but detectable stabilization of Arg-Ura3p by Arg-e$^{\Delta K}$-βgal alone (Fig. 5*C*) was consistent not only with the inability of Arg-e$^{\Delta K}$-βgal to confer the Ura^{+} phenotype on cells expressing Arg-Ura3p from uninduced P$_{CUP1}$ promoter but also with the partial rescue of the Ura^{+} phenotype by Arg-e$^{\Delta K}$-βgal in cells expressing Arg-Ura3p from the induced P$_{CUP1}$ (Figs. 3 and 4 and data not shown).

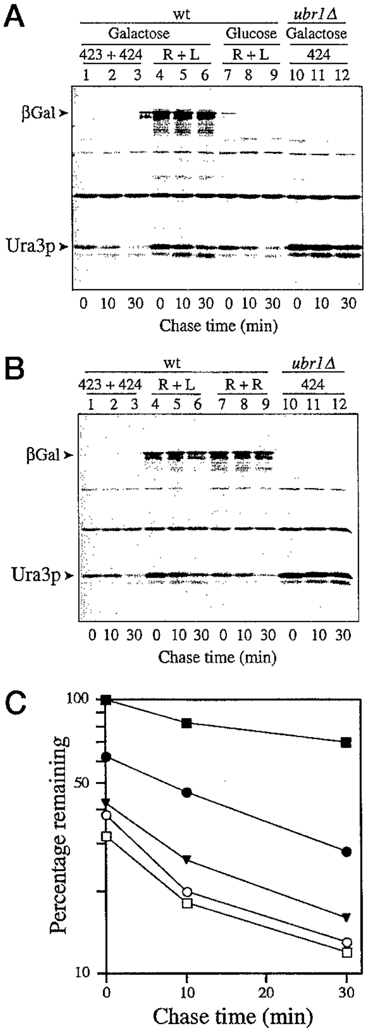


FIG. 5. **Metabolic stabilization of Arg-Ura3p in the presence of bivalent N-end rule inhibitor.** *A*, *S. cerevisiae* JD52 (*UBR1*) expressing Arg-Ura3p, a short-lived Ura3p-based reporter (Fig. 2*B*), from the induced P$_{CUP1}$ promoter, were cotransformed, alternatively, with either pRS423 (*HIS3*-based control vector) and pRS424 (P$_{GAL1}$, *TRP1*-based control vector) (denoted as *423+424*), or with pRΔ-βgal-TRP1 and pLΔ-βgal-HIS3, expressing Arg-e$^{\Delta K}$-βgal and Leu-e$^{\Delta K}$-βgal (denoted as *R+L*; the bivalent inhibitor). Control JD55 (*ubr1Δ*) cells expressing Arg-Ura3p were transformed with pRS424. Cells grown in either dextrose-containing SD medium (no expression of βgal) or galactose-containing SG medium were labeled with [^{35}S]methionine/cysteine for 5 min at 30 °C, followed by a chase for 0, 10, and 30 min, extraction, immunoprecipitation, and SDS-10% polyacrylamide gel electrophoresis. *B*, same as in *A*, but cells were also cotransformed with the plasmids pRΔ-βgal-TRP1 and pRΔ-βgal-HIS3, both expressing Arg-eΔK-βgal (denoted as *R+R*). The assays were carried out in SG medium. *C*, *in vivo* decay curves of Arg-Ura3p (Fig. 2*B*) in wild type (JD52) and *ubr1Δ* (JD55) cells. The patterns in *A* and *B* were quantified as described under "Experimental Procedures." The initial amounts of Arg-Ura3p were normalized against the amount in *ubr1Δ* cells (100%). Note that the inhibitors also altered the zero point effect (degradation of a reporter *during* the pulse (16, 42)). ●, degradation of Arg-Ura3p in *UBR1* (JD52) cells in the presence of both Arg-e$^{\Delta K}$-βgal and Leu-e$^{\Delta K}$-βgal (cells were grown in SG medium); ○, the same transformants were grown in SD medium where βgal fusions were not expressed; ▼, JD52 cells expressing Arg-Ura3p were transformed with the two alternatively marked plasmids expressing Arg-e$^{\Delta K}$-βgal and grown in SG medium; □, JD52 cells expressing Arg-Ura3p were transformed with the two alternatively marked control vectors and grown in SG medium; ■, *ubr1Δ* (JD55) cells expressing Arg-Ura3p were transformed with control vectors and grown in SG medium.

The Arg/Leu-e$^{\Delta K}$-βgal-based bivalent inhibitor of the present work, although surprisingly potent (Fig. 4*B*), is obviously far from optimal even for a protein-based inhibitor; because βgal is a homotetramer, only ~50% of the coexpressed Arg-e$^{\Delta K}$-βgal and Leu-e$^{\Delta K}$-βgal chains would exist as heterodimers within tetramers (Fig. 1*B*). (This estimate assumes a random assortment of Arg- and Leu-bearing βgal chains in the formation of βgal tetramers. The actual *in vivo* assortment is expected to be biased, to an unknown extent, in favor of homodimeric associations, because individual polysomes would produce βgal chains bearing either Arg or Leu but not both.) In addition, although the e$^{\Delta K}$ extension (Fig. 2*A*) is capable of supporting the desired effects, it is also unlikely to be optimal. In summary, the efficacy of this first and necessarily suboptimal bivalent inhibitor bodes well for the future of this design.

A bivalent inhibitor is strikingly more efficacious than an otherwise identical monovalent inhibitor (Figs. 3–5). In addition, our findings are the first evidence that the type 1 and type 2 sites of N-recognin are spatially proximal in the 225-kDa *S. cerevisiae* Ubr1p. While this work was under way, genetic dissection of *S. cerevisiae* Ubr1p identified amino acid residues that are required for the integrity of the type 1 site but not the type 2 site, and *vice versa*.[6] These results provided independent evidence for both the separateness and spatial proximity of the two substrate-binding sites of the 225-kDa N-recognin, in agreement with the present data. Our results (Figs. 3–5) strongly suggest that small bivalent inhibitors of the N-end rule pathway are feasible, and moreover, are expected to be much more potent than their monovalent counterparts. Work to produce such inhibitors is under way.

Acknowledgments—We thank members of the Varshavsky laboratory, especially A. Kashina and G. Turner, for helpful discussions and advice in the course of this work. We also thank I. Davydov, F. Du, F. Navarro-Garcia, H. Rao, and especially G. Turner for comments on the manuscript.

REFERENCES

1. Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 12142–12149
2. Bachmair, A., Finley, D., and Varshavsky, A. (1986) *Science* **234,** 179–186
3. Hershko, A. (1991) *Trends Biochem. Sci.* **16,** 265–268
4. Peters, J.-M., King, R. W., and Deshaies, R. J. (1998) in *Ubiquitin and the Biology of the Cell*, pp. 345–387, Plenum Press, New York
5. Varshavsky, A. (1997) *Trends Biochem. Sci.* **22,** 383–387
6. Hochstrasser, M. (1996) *Annu. Rev. Genet.* **30,** 405–439
7. Baumeister, W., Walz, J., Zühl, F., and Seemüller, E. (1998) *Cell* **92,** 367–380
8. Rechsteiner, M. (1998) in *Ubiquitin and the Biology of the Cell* (Peters, J. M., Harris, J. R., and Finley, D., eds) pp. 147–189, Plenum Press, New York
9. Kwon, Y. T., Kashina, A. S., and Varshavsky, A. (1999) *Mol. Cell. Biol.* **19,** 182–193
10. Grigoryev, S., Stewart, A. E., Kwon, Y. T., Arfin, S. M., Bradshaw, R. A., Jenkins, N. A., Copeland, N. G., and Varshavsky, A. (1996) *J. Biol. Chem.* **271,** 28521–28532
11. Stewart, A. E., Arfin, S. M., and Bradshaw, R. A. (1995) *J. Biol. Chem.* **270,** 25–28
12. Gonda, D. K., Bachmair, A., Wünning, I., Tobias, J. W., Lane, W. S., and Varshavsky, A. (1989) *J. Biol. Chem.* **264,** 16700–16712
13. Davydov, I. V., Patra, D., and Varshavsky, A. (1998) *Arch. Biochem. Biophys.* **357,** 317–325
14. Bartel, B., Wünning, I., and Varshavsky, A. (1990) *EMBO J.* **9,** 3179–3189
15. Kwon, Y. T., Reiss, Y., Fried, V. A., Hershko, A., Yoon, J. K., Gonda, D. K., Sangan, P., Copeland, N. G., Jenkins, N. A., and Varshavsky, A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 7898–7903
16. Baker, R. T., and Varshavsky, A. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 2374–2378
17. Reiss, Y., Kaim, D., and Hershko, A. (1988) *J. Biol. Chem.* **263,** 2693–2269
18. Byrd, C., Turner, G. C., and Varshavsky, A. (1998) *EMBO J.* **17,** 269–277
19. Madura, K., and Varshavsky, A. (1994) *Science* **265,** 1454–1458
20. Schauber, C., Chen, L., Tongaonkar, P., Vega, I., and Madura, K. (1998) *Genes Cells* **3,** 307–319
21. Alagramam, K., Naider, F., and Becker, J. M. (1995) *Mol. Microbiol.* **15,** 225–234
22. Ota, I. M., and Varshavsky, A. (1993) *Science* **262,** 566–569
23. deGroot, R. J., Rümenapf, T., Kuhn, R. J., and Strauss, J. H. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 8967–8971
24. Sijts, A. J., Pilip, I., and Pamer, E. G. (1997) *J. Biol. Chem.* **272,** 19261–19268
25. Hondermarck, H., Sy, J., Bradshaw, R. A., and Arfin, S. M. (1992) *Biochem. Biophys. Res. Commun.* **30,** 280–288
26. Taban, C. H., Hondermarck, H., Bradshaw, R. A., and Boilly, B. (1996) *Experientia* **52,** 865–870
27. Solomon, V., Lecker, S. H., and Goldberg, A. L. (1998) *J. Biol. Chem.* **273,** 25216–25222
28. Solomon, V., Baracos, V., Sarraf, P., and Goldberg, A. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 12602–12607
29. Tobias, J. W., Shrader, T. E., Rocap, G., and Varshavsky, A. (1991) *Science* **254,** 1374–1377
30. Varshavsky, A., Byrd, C., Davydov, I. V., Dohmen, R. J., Du, F., Ghislain, M., Gonzalez, M., Grigoryev, S., Johnson, E. S., Johnsson, N., Johnston, J. A., Kwon, Y. T., Lévy, F., Lomovskaya, O., Madura, K., Ota, I., Rümenapf, T., Shrader, T. E., Suzuki, T., Turner, G., Waller, P. R. H., and Webster, A. (1998) in *Ubiquitin and the Biology of the Cell* (Peters, J.-M., Harris, J. R., and Finley, D., eds) pp. 223–278, Plenum Press, New York
31. Kramer, R. H., and Karpen, J. W. (1998) *Nature* **395,** 710–713
32. Ptashne, M. (1992) *A Genetic Switch*, Cell Press, Cambridge, MA
33. Madura, K., Dohmen, R. J., and Varshavsky, A. (1993) *J. Biol. Chem.* **268,** 12046–12054
34. Sherman, F. (1991) *Methods Enzymol.* **194,** 3–21
35. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K. (eds) (1996) *Current Protocols in Molecular Biology*, Wiley-Interscience, New York
36. Bachmair, A., and Varshavsky, A. (1989) *Cell* **56,** 1019–1032
37. Johnson, E. S., Gonda, D. K., and Varshavsky, A. (1990) *Nature* **346,** 287–291
38. Sikorski, R. S., and Hieter, P. (1989) *Genetics* **122,** 19–27
39. Johnson, E. S., Ma, P. C., Ota, I. M., and Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 17442–17456
40. Jacobson, R. H., Zhang, X. J., DuBose, R. F., and Matthews, B. W. (1994) *Nature* **369,** 761–766
41. Ghislain, M., Dohmen, R. J., Levy, F., and Varshavsky, A. (1996) *EMBO J.* **15,** 4884–4899
42. Lévy, F., Johnsson, N., Rümenapf, T., and Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 4907–4912

---

[6] A. Webster, M. Ghislain, and A. Varshavsky, unpublished data.

Eur. J. Biochem. **259**, 244–252 (1999) © FEBS 1999

Analysis of a conditional degradation signal in yeast and mammalian cells

Frédéric Lévy*, Jennifer A. Johnston† and Alexander Varshavsky

Division of Biology, California Institute of Technology, Pasadena, CA, USA

The N-end rule pathway is a ubiquitin-dependent proteolytic system, the targets of which include proteins that bear destabilizing N-terminal residues. The latter are a part of the degradation signal called the N-degron. Arg-DHFR^{ts}, an engineered N-end rule substrate, bears N-terminal arginine (a destabilizing residue) and DHFR^{ts} [a temperature-sensitive mouse dihydrofolate reductase (DHFR) moiety]. Previous work has shown that Arg-DHFR^{ts} is long-lived at 23 °C but short-lived at 37 °C in the yeast *Saccharomyces cerevisiae*. In the present work, we extended this analysis, and found that the degradation of Arg-DHFR^{ts} can be nearly completely inhibited *in vivo* by methotrexate (MTX), a low-M_r ligand of DHFR. In *S. cerevisiae*, Arg-DHFR^{ts} is degraded at 37 °C exclusively by the N-end rule pathway, whereas in mouse cells the same protein at the same temperature is also targeted by another proteolytic system, through a degron in the conformationally perturbed DHFR^{ts} moiety. In mouse cells, MTX completely inhibits the degradation of Arg-DHFR^{ts} through its degron within the DHFR^{ts} moiety, but only partially inhibits degradation through the N-degron. When the N-terminus of Arg-DHFR^{ts} was extended with a 42-residue lysine-lacking extension, termed e$^{\Delta K}$, the resulting Arg-e$^{\Delta K}$-DHFR^{ts} was rapidly degraded at both 23 °C and 37 °C. Moreover, the degradation of Arg-e$^{\Delta K}$-DHFR^{ts}, in contrast with that of Arg-DHFR^{ts}, could not be inhibited by MTX, suggesting that the metabolic stability of Arg-DHFR^{ts} at 23 °C results, at least in part, from steric inaccessibility of its N-terminal arginine. The N-degron of Arg-DHFR^{ts} is the first example of a portable degradation signal the activity of which can be modulated *in vivo* by a cell-penetrating compound. We discuss implications of this advance and the mechanics of targeting by the ubiquitin system.

Keywords: degron; methotrexate; N-end rule; proteolysis; ubiquitin.

A number of regulatory circuits, including those that control the cell cycle, cell differentiation and responses to stress, involve metabolically unstable proteins [1–6]. A short *in vivo* half-life of a regulator provides a way to generate its spatial gradient and allows rapid adjustments of its concentration, or subunit composition, through changes in the rate of its synthesis. A protein can also be conditionally unstable, i.e. long-lived or short-lived, depending on the state of the cell. One example of the latter class are cyclins, a family of related proteins the destruction of which at specific stages of the cell cycle regulates cell division and growth [3,7].

Features of proteins that confer metabolic instability are called degradation signals, or degrons [8]. An essential component of one degradation signal, called the N-degron, is a destabilizing N-terminal residue of a protein [9,10]. A set of N-degrons containing different N-terminal residues which are destabilizing in a given cell yields a rule, termed the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. In eukaryotes, the N-end rule pathway is a part of the ubiquitin (Ub) system. Ub is a 76-residue protein whose covalent conjugation to other proteins plays a role in a multitude of processes, including cell growth, differentiation and responses to stress [2,4,11–13]. In many of these settings, Ub acts through routes that involve the degradation of Ub–protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease [14,15].

In eukaryotes, linear Ub–protein fusions are rapidly cleaved by Ub-specific proteases (UBPs) at the Ub–protein junction, making possible the production of otherwise identical proteins bearing different N-terminal residues, a technical advance that led to the discovery of the N-end rule [9,10]. When mouse dihydrofolate reductase (DHFR), a 20-kDa monomeric protein, was expressed as a Ub-Arg-DHFR fusion in the yeast *Saccharomyces cerevisiae* at 30 °C, the resulting Arg-DHFR (produced through deubiquitylation) was long-lived ($t_{0.5} > 4$ h), even though it bore N-terminal Arg, a destabilizing residue [16]. By contrast, a modified protein, Arg-e^K-DHFR (produced from Ub-Arg-e^K-DHFR), which bore a 42-residue, lysine-containing extension (denoted as e^K) between the N-terminal Arg residue and the first (Val) residue of DHFR, was short-lived *in vivo* ($t_{0.5}$ of ≈ 10 min), being degraded by the N-end rule pathway [16].

This and other evidence indicated that the N-degron comprises two essential determinants: a destabilizing N-terminal residue and an internal Lys residue (or residues) of a substrate. The Lys residue is the site of attachment of a multi-Ub chain [16–18]. Arg-e^K-DHFR contains a complete N-degron because it bears both an N-terminal destabilizing residue (Arg) and, in the e^K extension, at least one targetable Lys residue. In contrast, the N-degron of Arg-DHFR is incomplete, for at least one of the

Correspondence to Alexander Varshavsky, Division of Biology, 147–75, Caltech, 1200 East California Boulevard, Pasadena, CA 91125, USA.
Tel.: +1-626-395-3785, Fax: +1-626-440-9821,
E-mail: avarsh@cco.caltech.edu
**Present address*: Ludwig Institute for Cancer Research, Lausanne Branch, Ch. des Boveresses 155, CH-1066 Epalinges, Switzerland.
†*Present address*: Department of Biological Sciences, Stanford University, CA 94305, USA.

Abbreviations: βgal, *Escherichia coli* β-galactosidase; DHFR, mouse dihydrofolate reductase; ID, initial decay; MTX, methotrexate; Ub, ubiquitin; UBP, Ubspecific protease; UPR, ubiquitinproteinreference.

Note: F. Lévy and J. A. Johnston contributed equally to the present work.

(Received 28 July 1998; revised 12 October 1998; accepted 13 October 1998)

two mutually non-exclusive reasons: the absence of Lys residues accessible to the targeting complex of the N-end rule pathway and/or poor accessibility of the N-terminal Arg in Arg-DHFR to the same targeting complex. Since the 20-kDa mouse DHFR contains 16 Lys residues, this interpretation presumes that, in a folded DHFR molecule, the lysines are ineffective as ubiquitylation sites, because of the lysine's lack of mobility and/or its distance from a destabilizing N-terminal residue [10].

Previous work [19] described a temperature-sensitive (*ts*) allele of DHFR that was used to construct a heat-inducible N-degron, in a Ub-Arg-DHFR^{ts}-Ura3p fusion which contained *S. cerevisiae* Ura3p as a C-terminal reporter. The Ub-Arg-DHFR^{ts} moiety of this fusion was identical with Ub-Arg-DHFR above, except for a Pro → Leu alteration at position 66 of the DHFR moiety. Arg-DHFR^{ts}-Ura3p was long-lived at 23 °C but short-lived at 37 °C in yeast. Moreover, the Ub-Arg-DHFR^{ts} module of Ub-Arg-DHFR^{ts}-Ura3p was shown to be portable, in that linking it to a protein of interest conferred heat-inducible metabolic instability on the entire fusion [19].

In addition to yielding a new method for the construction of *ts* mutants, the heat-inducible N-degron [19] provided an approach to analyzing the mechanism of targeting by the N-end rule pathway. In carrying out this analysis, described below, we have found that the degradation of Arg-DHFR^{ts} can be inhibited *in vivo* by methotrexate (MTX), a low-M_r ligand of DHFR. We also show that, in mouse cells, Arg-DHFR^{ts} is targeted by both the N-end rule pathway and another proteolytic system, whereas in yeast the same test protein is exclusively an N-end rule substrate. The N-degron of Arg-DHFR^{ts} is the first example of a portable degradation signal whose activity can be modulated *in vivo* by a cell-penetrating compound.

MATERIALS AND METHODS

Plasmids for expression in *S. cerevisiae*

The plasmid pJJRtd, which expressed Ub-Arg-DHFR^{ts}-ha (Fig. 1, construct I), was based on the vector pRS316 [20]. A 1.0-kb *Eco*RI/*Hin*dIII fragment of pPW58 [8] was ligated into pRS316, followed by the cloning of the P_{CUP1} promoter-containing 0.4 kb *Eco*RI/*Sac*I fragment of pPW48 [19] into the *Eco*RI site, a step that yielded pJJRtd. pJJMtd, which expressed Ub-Met-DHFR^{ts}-ha, was prepared by ligating the large *Age*I/*Afl*II fragment of pJJRtd to a small fragment from the *Age*I/*Afl*II digest of pLGUbM-DHFRha [21].

Plasmids for expression in mammalian cells

Mouse DHFR and DHFR^{ts} open reading frames bearing the desired restriction sites at their 5′ and 3′ ends were produced using PCR. All DHFR and DHFR^{ts} moieties were flanked, at their C-termini, with the ha epitope [22] (Fig. 1), and were cloned into the pRc/CMV vector (Invitrogen, San Diego, CA, USA).

Constructs II and III (Fig. 1) were produced by PCR amplification of a fragment encoding Met-DHFR^{ts} and Arg-DHFR^{ts}, respectively, using construct I as a template. The N-terminal Met of construct II was specified by the PCR primer

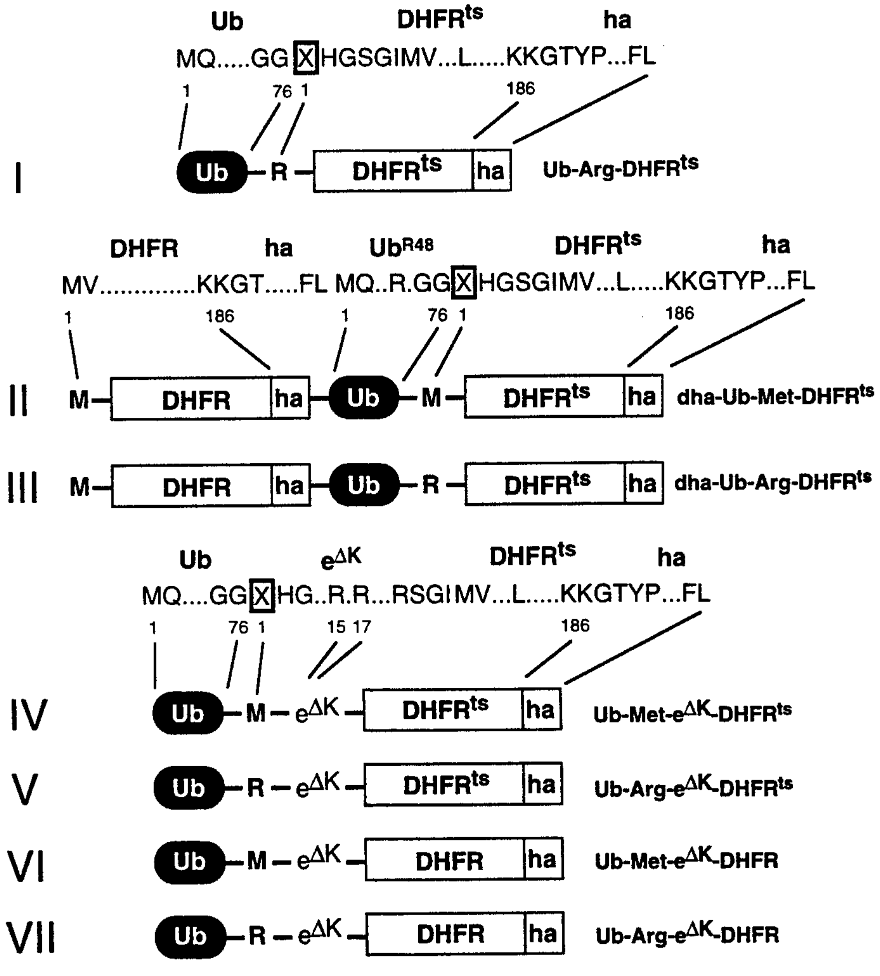


Fig. 1. Test proteins. Fusions used in this work contained some of the following moieties: the Ub moiety, either N-terminal (constructs I, IV–VII) or placed between other moieties (constructs II and III); an Arg residue at the junction between Ub and the C-terminal part of the fusion (constructs I, III, V and VII); a Met residue at the same junction (constructs II, IV, and VI); a 42-residue *E. coli* Lac repressor-derived sequence, termed e$^{\Delta K}$ [extension (e) lacking lysines (ΔK)], between Ub and the reporter part of the fusion (see the main text and [22,27]) (constructs IV–VII); DHFRha, a mouse DHFR moiety extended at the C-terminus by a sequence containing the hemagglutinin-derived *ha* epitope (constructs I–VII); DHFR^{ts}, a mutant temperature-sensitive DHFR moiety that differs from DHFR by the Pro → Leu alteration at position 66 [19] (see the main text). Amino acid residues are indicated by their single-letter abbreviations. Residues that vary between constructs are boxed in the sequences above the diagrams.

annealing to the 5′ end of the fragment. The PCR products were digested with *Xba*I, blunted with Klenow Pol I, followed by a cut with *Sac*II. A fragment bearing an open reading frame was inserted between the blunt-ended *Not*I site and the *Sac*II site of the plasmid pRc/dhaUbMbgal [23], resulting in the plasmids pRc/dha-Ub-Met-DHFR^{ts} and pRc/dha-Ub-Arg-DHFR^{ts}. Met- and Arg-DHFR^{ts} contained the six-residue sequence His-Gly-Ser-Gly-Ile-Met between the N-terminal residue and Val, the first natural residue of DHFR (Fig. 1).

To produce constructs IV–VII (Fig. 1), a fragment encoding Arg-DHFR^{ts} was amplified by PCR, using construct I as template. The PCR product was digested with *Xba*I, and blunted with Klenow Pol I, followed by a *Sac*II cut. The resulting fragment was inserted between the blunt-ended *Not*I site and the *Sac*II site of pRc/UbSβgal, yielding pRc/Ub-Arg-DHFR^{ts}. The latter served as a vector for the constructs IV–VII. Constructs IV and V were obtained by amplifying fragments encoding, respectively, Met-e$^{\Delta K}$-DHFR^{ts} and Arg-e$^{\Delta K}$-DHFR^{ts}, using the plasmid pLG/Ub^{V76}-Val-e$^{\Delta K}$DHFRha [22] as a template. The N-terminal residues Met and Arg were specified by a PCR primer annealing to the 5′ end of the fragment. Digestion of the PCR products with *Sac*II and *Pfl*MI yielded a fragment coding for Met- or Arg-e$^{\Delta K}$ and the first 39 residues of DHFR (DHFR$_{1-39}$). (*Pfl*MI cleaves the DHFR^{ts} open reading frame upstream of the mutated codon at position 66.) This fragment was inserted in-frame between the *Sac*II/*Pfl*MI sites of the plasmid pRc/Ub-Arg-DHFR^{ts}, yielding plasmids expressing Ub-Met-e$^{\Delta K}$-DHFR^{ts} and Ub-Arg-e$^{\Delta K}$-DHFR^{ts}. Constructs VI and VII were produced by exchanging a *SnaBI*/*Pfl*mI fragment of pRc/dhaUbNe^{K}DHFRha with the *SnaBI*/*Pfl*MI fragments derived from constructs IV and V, respectively. This step eliminated the fragment encoding dha-Ub-Asn-e^{K}-DHFR$_{1-13}$ of the original plasmid, replacing it, in-frame, with the sequence encoding either Ub-Met-e$^{\Delta K}$-DHFR$_{1-13}$ or Ub-Arg-e$^{\Delta K}$-DHFR$_{1-13}$. (The plasmid pRc/dhaUbNe^{K}DHFRha was constructed by replacing the region encoding βgal in pRc/dhaUbXβgal [23] with the region encoding DHFR.)

Yeast and mouse cell cultures, transfection, and pulse–chase analysis

S. cerevisiae were grown and manipulated as described previously [24] and in the legend to Fig. 2. Mouse L cells, a fibroblast-like cell line (ATCC CCL 1.3, American Type

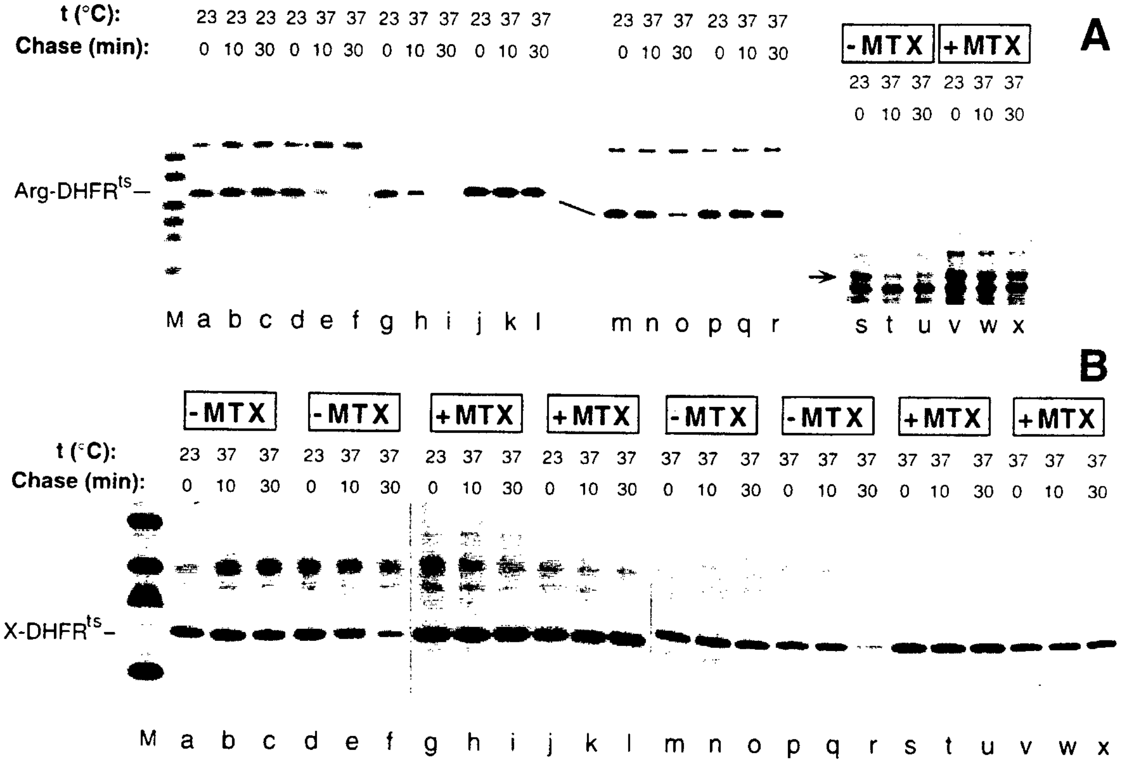


Fig. 2. Inhibition of Arg-DHFR^{ts} degradation by MTX in S. cerevisiae. (A) Lane M, molecular mass markers. Lanes a–c, JD47-13c (*UBR1*) cells expressing Arg-DHFR^{ts} (Ub-Arg-DHFR^{ts}) were labeled for 4 min at 23 °C with [^{35}S]methionine/cysteine, followed by a chase for 10 and 30 min as described [40], then extraction, immunoprecipitation, and Tricine-based SDS/PAGE (12% gel) of Arg-DHFR^{ts} [41]. Lanes d–f, same as lanes a–c, except that immediately after the 4-min pulse cells were shifted to 37 °C, and the chase was carried out at 37 °C. Lanes g–i, same as lanes d–f, but an independent pulse–chase experiment. Lanes j–l, same as lanes g–i, but with JD55 (*ubr1Δ*) *S. cerevisiae*. Lanes m–o, same as lanes g–i, but with KMY613 (*ubc2Δ*) *S. cerevisiae* strain expressing the plasmid-borne *UBC2* gene [42]. Lanes p–r, same as lanes m–o, but with KMY613 (*ubc2Δ*) lacking the plasmid-borne *UBC2*. Lanes s–u, JD47-13c (*UBR1*) cells expressing Arg-DHFR^{ts}-ha-Cdc28p (Ub-Arg-DHFR^{ts}-ha-Cdc28p) [19] were labeled and processed as described for lanes d–f. Lanes v–x, same as lanes s–u, but with MTX added to the growth medium to a final concentration of 20 μM 30 min before the pulse. (B) Lane M, molecular mass markers. Lanes a–c, JD47-13c (*UBR1*) cells expressing Met-DHFR^{ts} (Ub-Met-DHFR^{ts}) were labeled for 4 min at 23 °C with [^{35}S]methionine/cysteine. Immediately thereafter, the cells were shifted to 37 °C, followed by a chase at 37 °C for 10 and 30 min, extraction, immunoprecipitation, and SDS/PAGE of Met-DHFR^{ts}. Lanes d–f, same as lanes a–c, but with Arg-DHFR^{ts}. Lanes g–i, same as lanes a–c, but with MTX added to the growth medium 30 min before the pulse to a final concentration of 20 μM. Lanes j–l, same as lanes g–i, but with Arg-DHFR^{ts}. Lanes m–o, same as lanes a–c, except that both the pulse and the chase were at 37 °C. Lanes p–r, same as lanes m–o, but with Arg-DHFR^{ts}. Lanes s–u, same as lanes m–o, but with MTX added to the growth medium 30 min before the pulse to a final concentration of 20 μM. Lanes v–x, same as lanes s–u, but with Arg-DHFR^{ts}.

Culture Collection, Rockville, MD, USA), were grown as monolayers in Dulbecco's modified Eagle's medium/F12 medium, supplemented with 10% fetal bovine serum, antibiotics, 2 mM L-glutamine and 20 mM Hepes (sodium salt; pH 7.3). The cultures were regularly checked for the absence of mycoplasmas. Transient transfections and pulse–chase analyses were performed as described [23]. For the pulse–chases at 23 °C, cells were transferred to this temperature 30 min before labeling for 10 min with 0.1 mCi (1 Ci = 37 GBq) of [^{35}S]EXPRESS (New England Nuclear, Boston, MA, USA). Cells were chased at either 23 °C or 37 °C in the complete medium containing 0.1 mg·mL^{-1} cycloheximide. All media were pre-equilibrated at the desired temperature. In experiments that involved MTX, the drug was added 30 min before labeling and was present at a final concentration of 20 µM (diluted from 20 mM MTX in 0.9 M potassium/sodium phosphate, pH 7.3). The samples were analyzed by SDS/PAGE (15% gel), followed by autoradiography and quantitation using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA, USA).

To distinguish between partial half-lives that approximate the slopes of different regions of a non-exponential decay curve, a generalized half-life term, $t_{0.5}^{y-z}$, was used, where 0.5 denotes the parameter's half-life aspect and $y-z$ denotes the relevant time interval, from y to z min of chase. Another term [23], initial decay (ID), expressed as a percentage, equals 1 minus the ratio of the amount of a radiolabeled X-DHFR test protein (normalized against the reference protein, Met-DHFR-Ub; Fig. 1) at the end of the pulse (time 0) to the (normalized) amount of a radiolabeled long-lived X-DHFR such as Met-DHFR. As ID may depend on the duration of a pulse, a superscript, in ID^{x}, invokes the pulse time explicitly [23].

RESULTS

Inhibition of Arg-DHFR^{ts} degradation by MTX in *S. cerevisiae*

The major test protein of this work was Arg-DHFR^{ts} (produced from Ub-Arg-DHFR^{ts}) (Fig. 1, construct I), identical with the Arg-DHFR^{ts} moiety of a larger fusion described in the introduction. The test proteins bore C-terminal *ha*, a hemagglutinin-derived epitope tag recognized by a monoclonal antibody [22] (Fig. 1; the *ha* epitope is omitted below in the names of specific test proteins). For the analysis in *S. cerevisiae*, Arg-DHFR^{ts} was expressed from the copper-inducible P_{CUP1} promoter in a low-copy vector, and the metabolic fate of Arg-DHFR^{ts} was determined in a pulse–chase assay (see Materials and methods). Similarly to the previously examined Arg-DHFR^{ts}-Ura3p and Arg-DHFR^{ts}-Cdc28p [19], Arg-DHFR^{ts} was a long-lived protein at 23 °C ($t_{0.5}$ > 4 h) but a short-lived protein at 37 °C ($t_{0.5} \approx$ 4 min) (Fig. 2A, lanes a–c vs. d–f). The degradation of Arg-DHFR^{ts} at 37 °C was carried out by the N-end rule pathway, as indicated by (a) the requirement for Ubr1p (E3), the recognition component of this pathway (Fig. 2A, lanes g–i vs. j–l), (b) the requirement for Ubc2p (E2), the relevant Ub-conjugating enzyme (Fig. 2A, lanes m–o vs. p–r), and (c) the long half-life of Met-DHFR^{ts}, an otherwise identical protein that bore N-terminal Met, a stabilizing residue, instead of Arg (Fig. 2B, lanes a–c vs. d–f).

Previous work [27] has shown that the degradation of Arg-e^{K}-DHFR by the N-end rule pathway in an extract from rabbit reticulocytes can be inhibited by the folate analog MTX, a high-affinity DHFR ligand ($K_d \approx$ 10 pM). It was far from clear whether MTX would have the same effect *in vivo*, in part because the purified N-end rule substrates added to the extract contained the folded DHFR moieties [27]. By contrast, a nascent DHFR-based substrate is unable to bind MTX until after the folding of DHFR [28], but can be targeted by the N-end rule pathway *in vivo* at any time, possibly even during translation [16].

The metabolic fate of Arg-DHFR^{ts} at 37 °C was monitored in *S. cerevisiae* in the absence or presence of 20 µM MTX in the medium. MTX was found to be an efficacious inhibitor of Arg-DHFR^{ts} degradation in *S. cerevisiae* (Fig. 2B, lanes d–f and p–r vs. j–l and v–x). MTX inhibited the degradation of Arg-DHFR^{ts} regardless of whether it was added to the growth medium 15 min or 24 h before the pulse–chase assay (data not shown; the growth of cells was not affected significantly by 20 µM MTX for at least 24 h). The protective effect of MTX on Arg-DHFR^{ts} was essentially the same regardless of whether the pulse labeling (followed by the 37 °C chase) was carried out at 23 °C (Fig. 2B, lanes d–f vs. j–l) or at 37 °C (Fig. 2B, lanes p–r vs. v–x).

The degradation (at 37 °C) of Arg-DHFR^{ts}-Cdc28p, which differed from Arg-DHFR^{ts} (produced from Ub-Arg-DHFR^{ts}) by the presence of the Cdc28p kinase moiety, was also inhibited by MTX (Fig. 2A, lanes s–u vs. v–x). Arg-DHFR^{ts}-Cdc28p, which was long-lived at 23 °C, or at 37 °C in *ubr1Δ* cells, could function as a Cdc28p kinase *in vivo* [19]. As Arg-DHFR^{ts}-Cdc28p was short-lived at 37 °C in *UBR1 S. cerevisiae* (Fig. 2A, lanes s–u), the cells that also carried a mutant (*ts*) *cdc28* allele [19] were not viable at 37 °C, owing to the absence of the essential Cdc28p kinase at this temperature. The addition of MTX (at 20 µM) was found to rescue these cells at 37 °C through the inhibition of degradation of Arg-DHFR^{ts}-Cdc28p (Fig. 2A, lanes s–u vs. v–x, and data not shown). The effect of MTX was specific for DHFR-containing substrates, as degradation of unrelated test proteins, for example Arg-βgal (Ub-Arg-βgal), by the N-end rule pathway was unimpaired in the presence of MTX (data not shown).

DHFR as an N-end rule substrate in mammalian cells

DHFR-based substrates of the N-end rule pathway were used in the earlier studies with *S. cerevisiae* [16,19], but have not been examined in mammalian cells. We used L cells, a mouse fibroblast-like cell line, and the ubiquitin/protein/reference (UPR) technique, which increases the accuracy of pulse–chase assays [26] by providing a reference protein [23]. In the UPR technique, Ub is located between a protein of interest and a reference protein in a linear fusion (e.g. construct II in Fig. 1). This fusion is cleaved, cotranslationally or nearly so, by UBPs at the last residue of Ub, producing equimolar amounts of the protein of interest and the reference protein bearing a C-terminal Ub moiety. If both the reference protein and the protein of interest are immunoprecipitated in a pulse–chase assay, the relative amounts of the protein of interest can be normalized against the reference in the same sample. The UPR technique can thus compensate for the scatter of immunoprecipitation yields, sample volumes and other sources of sample-to-sample variation [23].

The UPR constructs of the present work were fusions containing the metabolically stable Met-DHFR-Ub moiety as a reference protein (termed dha-Ub) and a DHFR-based N-end rule substrate such as, for example, Arg-DHFR^{ts} as a protein of interest (Fig. 1, construct III). To preclude the possibility that the C-terminal Ub moiety of dha-Ub could function as a ubiquitylation/degradation signal, the Lys48 residue of Ub (a major site of isopeptide bonds in multi-Ub chains [17,22,29]) was converted into Arg, which cannot be ubiquitylated [12].

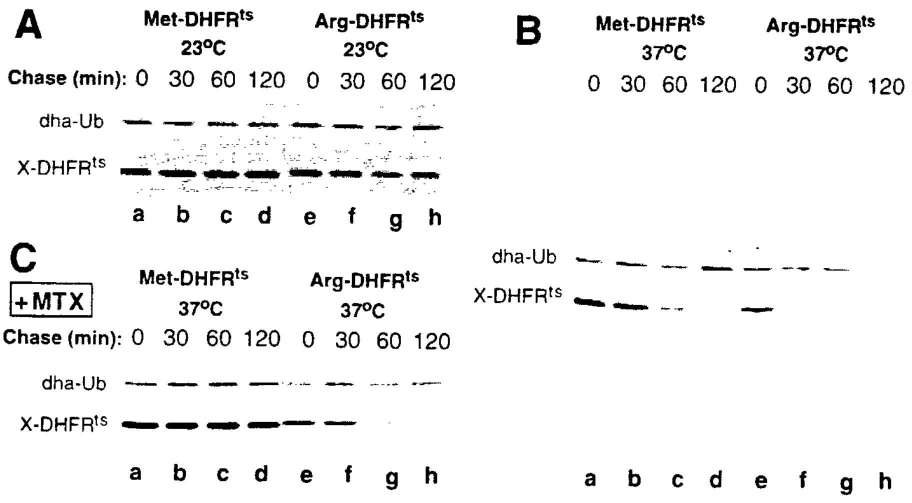


Fig. 3. Inhibition of Arg-DHFR^{ts} degradation in mouse cells by MTX. (A) Lanes a–d, mouse L cells transiently expressing a UPR-based version of Met-DHFR^{ts} (Met-DHFR-ha-Ub^{R48}-Met-DHFR^{ts}-ha; see Fig. 1, construct II) were labeled at 23 °C for 10 min with [^{35}S]methionine/cysteine, followed by a chase, also at 23 °C, for 30, 60 and 120 min in the presence of cycloheximide, extraction, immunoprecipitation, and SDS/PAGE of Met-DHFR^{ts}-ha and Met-DHFR-ha-Ub^{R48}. Lanes e–h, same as lanes a–d but with Arg-DHFR^{ts} (Met-DHFR-ha-Ub^{R48}-Arg-DHFR^{ts}-ha). (B) Lanes a–d, same as lanes a–d in (A), except that both the pulse and the chase were at 37 °C. Lanes e–h, same as lanes a–d, but with Arg-DHFR^{ts}-ha (Met-DHFR-ha-Ub^{R48}-Arg-DHFR^{ts}-ha). (C) Lanes a–d, same as lanes a–d in (B) but with 20 µM MTX in the medium. Lanes e–h, same as lanes a–d, but with Arg-DHFR^{ts} (Met-DHFR-ha-Ub^{R48}-Arg-DHFR^{ts}-ha). The bands of Arg/Met-DHFR^{ts}-ha and Met-DHFR-ha-Ub^{R48} (denoted as dha-Ub) are indicated. A larger area of the gel is shown for lanes a–h in (B), to illustrate the overall immunoprecipitation pattern.

These fusions were expressed from the cytomegalovirus early promoter, P_{CMV}, in mouse L cells. The cells were labeled for 10 min at either 23 °C or 37 °C with [^{35}S]methionine/cysteine, followed by a chase of 30, 60 and 120 min at the same temperature in the presence of cycloheximide, immunoprecipitation with anti-ha monoclonal antibody, and analysis/quantitation of immunoprecipitated proteins by SDS/PAGE and PhosphorImager, using UPR (see Materials and methods). Pulse–chase assays in which cycloheximide (a translation inhibitor) was omitted yielded similar results (data not shown).

Both Met-DHFR^{ts} and Arg-DHFR^{ts} were long-lived in mouse cells at 23 °C ($t_{0.5} > 10$ h) (Fig. 3A, lanes a–h, and Fig. 4A). By contrast, Arg-DHFR^{ts} was short-lived at 37 °C in mouse cells: its half-life, determined between 0 and 60 min of chase and denoted as $t_{0.5}^{0-60}$ [23], was ≈ 10 min (Fig. 3B, lanes e–h, and Fig. 4A). In addition, a large fraction (≈ 30%) of the pulse-labeled Arg-DHFR^{ts} (but not of Met-DHFR^{ts}) was degraded during the 10-min pulse at 37 °C, as could be seen from comparing the UPR-normalized amounts of Arg-DHFR^{ts} at time 0 (the beginning of the chase) at 23 °C and 37 °C (Fig. 4A). In contrast with a conventional pulse–chase assay, the use of UPR in a setting where a protein can be made either short-lived or long-lived allows the detection and measurement of proteolysis not only during the chase but during the pulse as well [23]. The extent of degradation of a protein during the pulse is denoted as ID^{t} [23]. In the present context, the variable ID^{t} is defined as the extent of degradation of a radiolabeled short-lived protein (Arg-DHFR^{ts} at 37 °C) at the end of a pulse relative to the amount of a nearly identical protein (Met-DHFR^{ts} at 37 °C) that is long-lived under the same conditions.

A large fraction, but not all, of the Arg-DHFR^{ts} degradation in mouse cells was carried out by the N-end rule pathway. This could be seen by comparing the UPR-based decay curves of Met-DHFR^{ts} and Arg-DHFR^{ts} at 37 °C (Fig. 3B lanes a–d vs. lanes e–h and Fig. 4A). Specifically, not only Arg-DHFR^{ts} but also Met-DHFR^{ts} were metabolically unstable at 37 °C in mouse cells, the corresponding $t_{0.5}^{0-60}$ being ≈ 10 min and ≈ 48 min, respectively (Fig. 4A). As Met is a stabilizing residue in the N-end rule [23,30], we infer that Arg-DHFR^{ts} was targeted not only by the N-end rule pathway, but also by another proteolytic system, which recognized a structural feature that resulted from a conformational perturbation of the DHFR^{ts} moiety at 37 °C. This aspect of Arg-DHFR^{ts} acted as a degron in mouse cells but not in yeast, inasmuch as the same protein was targeted exclusively by the N-end rule pathway in *S. cerevisiae* at 37 °C (Fig. 2A lanes g–i vs. lanes j–l). The nature of this degradation signal in the conformationally perturbed (at 37 °C) DHFR^{ts} moiety is not known.

Inhibition of both degradation signals of Arg-DHFR^{ts}-ha by MTX in mouse cells

Mouse L cells expressing Met-DHFR^{ts} (dha-Ub-Met-DHFR^{ts}) or Arg-DHFR^{ts} (dha-Ub-Arg-DHFR^{ts}) were incubated at 37 °C for 30 min with 20 µM MTX in the medium, followed by a 10-min pulse with [^{35}S]methionine/cysteine, a chase, and SDS/PAGE analysis. The degradation of Met-DHFR^{ts} at 37 °C, which is mediated by a non-N-degron in the conformationally perturbed DHFR^{ts} moiety (see above), was virtually completely inhibited by MTX (Fig. 3B lanes a–d vs. Fig. 3C lanes a–d and Fig. 4B). Specifically, the $t_{0.5}^{0-60}$ of Met-DHFR^{ts} at 37 °C was ≈ 48 min in the absence of MTX and > 10 h in the presence of MTX (Fig. 4B). The much more rapid degradation of Arg-DHFR^{ts} carried out by both the N-end rule pathway and the other proteolytic pathway was strongly but incompletely inhibited by MTX (Fig. 3B lanes e–h vs. Fig. 3C lanes e–h and Fig. 4B). Specifically, in the absence of MTX, the $t_{0.5}^{0-60}$ of Arg-DHFR^{ts} was ≈ 10 min (ID^{10} of 36%) (Fig. 4B). In the presence of MTX, these variables changed to $t_{0.5}^{0-60}$ ≈ 37 min and ID^{10} of 29% (Fig. 4B).

A comparison of the corresponding decay curves (Fig. 4), made more reliable by the increased accuracy of a UPR-based pulse–chase assay, suggested one reason for the leaky inhibition of the N-degron of Arg-DHFR^{ts} by MTX. Specifically, the ID^{10} of Met-DHFR^{ts} (i.e. the extent of Met-DHFR^{ts} degradation during the pulse) was < 10% in the absence of MTX, whereas

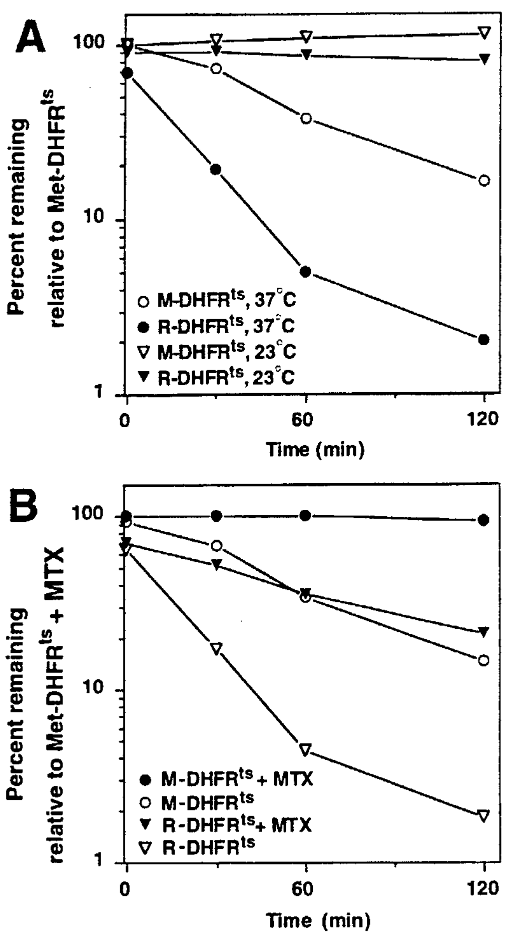


Fig. 4. Decay curves of Met-DHFR^{ts} and Arg-DHFR^{ts} at 23 °C and 37 °C in mouse cells, in the absence (A) and presence of MTX (B). These curves were derived from the UPR-based electrophoretic data (Figs 3 and 5 and analogous evidence), as described in Materials and methods.

the ID^{10} of Arg-DHFR^{ts} under the same conditions was ≈ 36%. In other words, more than one-third of labeled Arg-DHFR^{ts} molecules were degraded during the 10-min pulse (Fig. 4B). Moreover, the ID^{10} of Arg-DHFR^{ts} remained high (≈ 30%) even in the presence of MTX (Fig. 4B). This result suggested that a large fraction of the newly formed Arg-DHFR^{ts} molecules was targeted for degradation by the N-end rule pathway before the conformation of Arg-DHFR^{ts} was mature enough to allow high-affinity binding of MTX (see the Discussion).

An exposed N-terminus renders Arg-DHFR^{ts} constitutively short-lived

To address the relative contributions of the first (N-terminal Arg) and the second (internal Lys) determinant of the N-degron in Arg-DHFR^{ts}, we extended its N-terminus with a 42-residue, *Escherichia coli* Lac repressor-derived sequence, termed e$^{\Delta K}$ [extension (e) lacking lysines (ΔK)] [16]. Met-e$^{\Delta K}$-DHFR^{ts} (Ub-Met-e$^{\Delta K}$-DHFR^{ts}) (Fig. 1, construct IV), which bore a stabilizing N-terminal residue, was long-lived at 23 °C in mouse cells ($t_{0.5}^{0\text{-}60}$ > 10 h) (Fig. 5A, lanes a–c). Arg-e$^{\Delta K}$-DHFR^{ts} (Ub-Arg-e$^{\Delta K}$-DHFR^{ts}) (Fig. 1, construct V), which bore a destabilizing N-terminal residue, was short-lived at 23 °C ($t_{0.5}^{0\text{-}60}$ ≈ 45 min; Fig. 5A, lanes d–f), in contrast with the otherwise identical Arg-DHFR^{ts} that lacked the e$^{\Delta K}$ extension and was long-lived at 23 °C (Fig. 3A, lanes e–h). Note that Arg-e$^{\Delta K}$-DHFR^{ts} was short-lived at 23 °C, in spite of the fact that the e$^{\Delta K}$ extension did not contribute additional Lys residues to Arg-DHFR^{ts}, which was long-lived at 23 °C.

At 37 °C, Met-e$^{\Delta K}$-DHFR^{ts} was approximately as short-lived as Met-DHFR^{ts}; the degradation of these proteins was carried out by a non-N-end rule pathway that targeted a degron (d$_x$) in the conformationally perturbed DHFR^{ts} moiety (Fig. 5B, lanes a–c; compare with Fig. 3B, lanes a–d). In contrast, the degradation of Arg-e$^{\Delta K}$-DHFR^{ts} at 37 °C was much faster than that of Met-e$^{\Delta K}$-DHFR^{ts}, the difference in rate being due to the targeting of Arg-e$^{\Delta K}$-DHFR^{ts} by both the N-end rule pathway

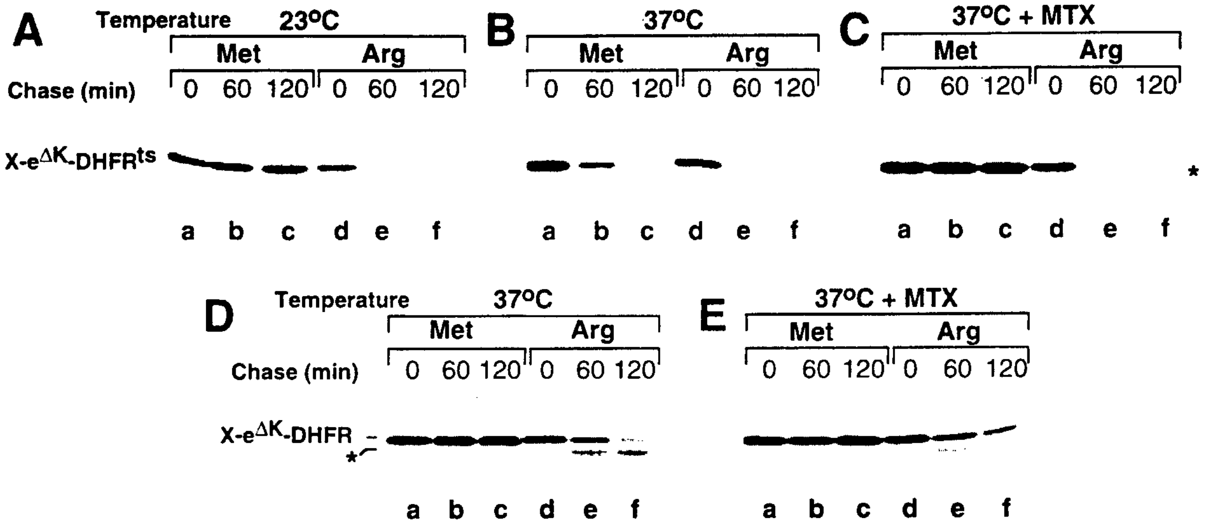


Fig. 5. Effects of temperature and MTX on the degradation of N-terminally extended DHFR^{ts} and DHFR in mouse cells. (A) Lanes a–c, mouse L-cells transiently expressing Met-e$^{\Delta K}$-DHFR^{ts} (Fig. 1, construct IV) were labeled at 23 °C for 10 min with [^{35}S]methionine/cysteine, followed by a chase, also at 23 °C, for 60 and 120 min in the presence of cycloheximide, extraction, immunoprecipitation, and SDS/PAGE analysis. Lanes d–f, same as lanes a–c but with Arg-e$^{\Delta K}$-DHFR^{ts}. (B) Lanes a–f, same as lanes a–f in (A), but the pulse–chase was carried out at 37 °C. (C) Lanes a–c, same as lanes a–c in (B), but the pulse–chase of Met-e$^{\Delta K}$-DHFR^{ts} was carried out in the presence of 20 µM MTX in the growth medium. Lanes d–f, same as lanes a–c, but with Arg-e$^{\Delta K}$-DHFR^{ts}. (D) Lanes a–c, same as lanes a–c in (B), but with Met-e$^{\Delta K}$-DHFR (the wild-type DHFR moiety). Lanes d–f, same as lanes a–c, but with Arg-e$^{\Delta K}$-DHFR. The asterisk indicates a cleavage product derived from a minority of short-lived Arg-e$^{\Delta K}$-DHFR^{ts} and Arg-e$^{\Delta K}$-DHFR molecules during their targeting by the N-end rule pathway (see the main text). (E) Lanes a–c, same as lane a–c in (D), but in the presence of 20 µM MTX. Lanes d–f, same as lanes a–c but with Arg-e$^{\Delta K}$-DHFR.

and the d_s-recognizing pathway (Fig. 5B, lanes d–f). A minor proteolytic fragment of Arg-$e^{\Delta K}$-$DHFR^{ts}$, visible in Fig. 5B–E (lanes d–f), was specific to N-end rule substrates bearing an $e^{\Delta K}$-type extension: an analogous fragment was also observed with X-e^{K}-βgal-based N-end rule substrates [16].

We also examined the metabolic fate of Met-$e^{\Delta K}$-DHFR (Ub-Met-$e^{\Delta K}$-DHFR) and Arg-$e^{\Delta K}$-DHFR (Ub-Arg-$e^{\Delta K}$-DHFR), which were identical with the substrates above except that they bore the wild-type mouse DHFR moiety (Fig. 1, constructs VI and VII). Met-$e^{\Delta K}$-DHFR was long-lived at both 23 °C and 37 °C in mouse cells (Fig. 5D, lanes a–c, and data not shown), indicating that the heat-activated d_s-type degron of substrates containing the $DHFR^{ts}$ moiety (e.g. Fig. 4A) resulted from the Pro66 gr; Leu66 mutation that yielded this moiety. The degradation of Arg-$e^{\Delta K}$-DHFR was mediated exclusively by the N-end rule pathway (Fig. 5D, lanes d–f vs. lanes a–c), unlike the double-pathway degradation of Arg-$e^{\Delta K}$-$DHFR^{ts}$ at 37 °C (Fig. 5C, lanes d–f). In contrast with the results with Arg-$DHFR^{ts}$ and Met-$DHFR^{ts}$, the presence of MTX did not result in a significant stabilization of Arg-$e^{\Delta K}$-DHFR (Fig. 5D lanes d–f vs. Figure 5E lanes d–f).

DISCUSSION

We report the following results.

(a) Arg-$DHFR^{ts}$ is long-lived at 23 °C but short-lived at 37 °C in both *S. cerevisiae* and mouse cells.

(b) In yeast, Arg-$DHFR^{ts}$ is degraded (at 37 °C) exclusively by the N-end rule pathway, whereas in mouse cells the same protein at the same temperature is degraded by another proteolytic pathway as well. The corresponding mouse-specific degron is active at 37 °C but inactive at 23 °C. This degradation signal is a currently unknown feature of the perturbed conformation of the $DHFR^{ts}$ moiety at 37 °C, which differs from the wild-type DHFR moiety by a Pro → Leu alteration at position 66.

(c) MTX, a low-M_r ligand of DHFR ($K_d \approx 10$ pM), inhibits the degradation of Arg-$DHFR^{ts}$ (at 37 °C) nearly completely in yeast and partially in mouse cells.

(d) The ability of MTX to inhibit the degradation of Arg-$DHFR^{ts}$ *in vivo* is retained when Arg-$DHFR^{ts}$ is utilized as a portable degron that confers metabolic instability on linked unrelated proteins in yeast.

(e) Using the UPR technique [23] to determine the metabolic fate of Arg-$DHFR^{ts}$ in mouse cells, we found that ≈ 36% of the labeled Arg-$DHFR^{ts}$ molecules are degraded during the 10-min pulse, and that the presence of MTX decreases this fraction only slightly, to ≈ 30%. This finding accounts for the leaky inhibition of Arg-$DHFR^{ts}$ degradation by MTX in mouse cells. Specifically, this finding suggests that a large fraction of the nascent Arg-$DHFR^{ts}$ is targeted, in mouse cells, by the N-end rule pathway before the conformation of the $DHFR^{ts}$ moiety is mature enough [28] to allow high-affinity binding of MTX. That the MTX-mediated inhibition of degradation of Arg-$DHFR^{ts}$ in yeast is much less leaky than in mouse cells is but one difference in the detailed properties of the yeast and mammalian N-end rule pathways [see also item (f)].

(f) When the N-terminus of Arg-$DHFR^{ts}$ was extended with a 42-residue lysine-lacking extension, termed $e^{\Delta K}$, the N-degron of the resulting Arg-$e^{\Delta K}$-$DHFR^{ts}$ was active at both 23 °C and 37 °C, unlike the mouse-specific degron, which remained inactive at 23 °C. Moreover, the degradation of Arg-$e^{\Delta K}$-$DHFR^{ts}$, in contrast with that of Arg-$DHFR^{ts}$, could not be inhibited by MTX, suggesting that the inactivity of the N-degron in Arg-$DHFR^{ts}$ at 23 °C results at least in part from inaccessibility of its N-terminal arginine to the targeting complex of the N-end rule pathway. Arg-$e^{\Delta K}$-DHFR, which bore the wild-type DHFR moiety, was also short-lived in mouse cells. Previously, the same protein was found to be long-lived in yeast, unless the $e^{\Delta K}$ extension was replaced by an otherwise identical extension, termed e^{K}, which contained Lys residues that functioned as the second determinant of the N-degron [16].

Potential applications of a heat-inducible, MTX-suppressible portable N-degron include its use to produce conditional mutants in homeothermic animals such as mouse. One difficulty in the current approaches to this problem [31–33] is that either a deletion or transcriptional repression of a gene of interest leaves the previously produced gene product unperturbed. If this protein is long-lived and resides in a non-dividing cell, there may be a considerable phenotypic lag between the conditionally introduced genetic change and the actual inactivation or disappearance of a corresponding protein. A degron that can be regulated by a cell-penetrating ligand can be employed to address this problem. Since chronic administration of MTX is toxic to mammals, it will be necessary to construct an analogous DHFR-based N-degron that uses *E. coli* DHFR and is inhibited by trimethoprim, which binds tightly to the *E. coli* but not the mouse DHFR [34]. If a portable N-degron of this kind could be constructed, combining it with a conditional repression of a gene of interest should yield better methods for producing conditional mutants in homeothermic animals.

MTX, through its binding to the substrate pocket of DHFR, stabilizes DHFR conformation [35,36]. This effect was employed in several studies with DHFR as a reporter protein. For example, the binding of MTX to DHFR can preclude, under certain conditions, the translocation of a DHFR-containing fusion protein across biological membranes [37,38]. MTX can partially protect DHFR against *in vitro* proteolysis by thermolysin [38,39]. MTX has also been shown to block the degradation of a targeted multiubiquitylated Arg-e^{K}-DHFR in reticulocyte extract [27]. In the latter study, the N-end rule pathway in reticulocyte extract was presented with a prefolded wild-type DHFR moiety. By contrast, in the present work with intact cells, the same pathway could target a DHFR-based N-end rule substrate immediately after (and possibly even during) its synthesis. Further analysis is required to clarify the apparent discrepancy between the previously observed inhibition of degradation of Arg-e^{K}-DHFR by MTX *in vitro* (in reticulocyte extract) [27] and the absence of a comparable effect of MTX *in vivo* on this class of substrates, in which N-terminal Arg is constitutively exposed. One difference between the two settings is the presentation of prefolded DHFR moieties in reticulocyte extract compared with a kinetic competition between the folding of newly formed DHFR moieties and their targeting for degradation by the N-end rule pathway.

The use of the UPR technique not only increased the overall accuracy of pulse–chase assays with mouse cells but also showed that ≈ 30% of the newly made Arg-$DHFR^{ts}$ was degraded during the pulse. The finding that the fraction of Arg-$DHFR^{ts}$ degraded during the pulse decreased only slightly in the presence of MTX strongly suggests that most of the labeled Arg-$DHFR^{ts}$ is conformationally immature shortly after synthesis, and therefore cannot form a high-affinity complex with MTX. The Pro → Leu alteration at position 66 that yielded $DHFR^{ts}$ occurred in the region of DHFR that interacts with the aromatic ring of MTX [35], and is likely to have resulted in a decreased affinity of DHFR for MTX. Nonetheless, the test proteins containing the $DHFR^{ts}$ moiety were efficiently retained on an MTX affinity column (data not shown), and their *in vivo* degradation could be specifically inhibited by MTX.

A model of targeting by the N-end rule pathway that accounts for our findings with the heat-inducible N-degron assumes that the main cause of metabolic stability of Arg-DHFR^{ts} at 23 °C is not the absence of a sterically accessible Lys residue within the folded DHFR globule, but rather a steric inaccessibility of the N-terminal Arg residue. In this model, the activation of the previously cryptic N-degron of Arg-DHFR^{ts} at 37 °C is caused by an increased exposure and flexibility of a region bearing the N-terminal Arg, a transition that leads to the binding of Arg by the targeting complex of the N-end rule pathway. According to this interpretation, MTX stabilizes Arg-DHFR^{ts} against degradation at 37 °C not by precluding the conformational mobilization of internal Lys residues, but by precluding a temperature-mediated increase in the exposure of the N-terminal Arg. In addition to being consistent with the available evidence, this model accounts for the otherwise puzzling result that Arg-e$^{\Delta K}$-DHFR^{ts}-ha (which differs from Arg-DHFR^{ts} by the presence of e$^{\Delta K}$) is short-lived at both 23 °C and 37 °C, whereas Arg-DHFR^{ts} is short-lived only at 37 °C. Indeed, the e$^{\Delta K}$ extension would result in a temperature-independent enhanced exposure of N-terminal Arg, a change that would be, in this model, sufficient for the temperature-independent activity of the N-degron in Arg-e$^{\Delta K}$-DHFR^{ts}. The earlier model proposed by Dohmen *et al.* [19], which presumed that the inactivity of the N-degron in Arg-DHFR^{ts} at 23 °C was caused by the absence of sterically accessible Lys residues, cannot account for the above result.

How can we verify the Arg-exposure model? One way is suggested by the existence, in both yeast and mammals, of the enzymes N-terminal amidase (Nt-amidase) and Arg-tRNA-protein transferase (R-transferase), which chemically modify specific (sterically exposed) N-terminal residues of proteins in the cytosol. In *S. cerevisiae*, the *NTA1*-encoded Nt-amidase deamidates N-terminal Asn or Gln; the *ATE1*-encoded R-transferase conjugates Arg to N-terminal Asp or Glu (reviewed in [10]). A test of the model would involve the expression of, for example, Asn-DHFR^{ts} (identical to Arg-DHFR^{ts} except for the presence of N-terminal Asn) in yeast cells that lack Ubr1p, a component of the N-end rule pathway that recognizes primary destabilizing N-terminal residues such as Arg. (The absence of Ubr1p would preclude the degradation of test proteins.) The model predicts that at 23 °C the N-terminal Asn of Asn-DHFR^{ts} would be deamidated inefficiently or not at all, in contrast with what happens at 37 °C. This can be verified by isolating Asn-DHFR^{ts} from cells incubated at either 23 °C or 37 °C, and determining its N-terminal residue. If the model is correct, the N-terminal sequence must begin largely or entirely with Asn at 23 °C, but would become Arg-Asp-... at 37 °C. (In the latter case, the N-terminal Asp, produced from Asn, would be arginylated by R-transferase.) Experimental verification of this model is under way.

ACKNOWLEDGEMENTS

We thank Michel Ghislain, Ailsa Webster and Martin Gonzalez for helpful discussions and advice, and Lawrence Peck for comments on the manuscript. This work was supported by the grants to A.V. from the National Institutes of Health (DK39520 and GM31530). F.L. and J.A.J. were supported by fellowships from the Swiss National Fund for Research and the American Cancer Society, respectively.

REFERENCES

1. Haas, A.J. & Siepman, T.J. (1997). Pathways of ubiquitin conjugation. *FASEB J.* **11**, 1257–1268.
2. Hershko, A. (1997). Roles of ubiquitin-mediated proteolysis in cell cycle control. *Curr. Op. Cell. Biol.* **9**, 788–799.
3. King, R.W., Deshaies, R.J., Peters, J.M. & Kirschner, M.W. (1996). How proteolysis drives the cell cycle. *Science* **274**, 1652–1659.
4. Varshavsky, A. (1997). The ubiquitin system. *Trends Biochem. Sci.* **22**, 383–387.
5. Wilkinson, K.D. (1997). Regulation of ubiquitin-dependent processes by deubiquitinating enzymes. *FASEB J.* **11**, 1245–1256.
6. Pickart, C.M. (1997). Targeting of substrates to the 26S proteasome. *FASEB J.* **11**, 1055–1066.
7. Murray, A. & Hunt, T. (1993) *The Cell Cycle*. Freeman, New York.
8. Varshavsky, A. (1991). Naming a targeting signal. *Cell* **64**, 13–15.
9. Bachmair, A., Finley, D. & Varshavsky, A. (1986). *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179–186.
10. Varshavsky, A. (1997). The N-end rule pathway of protein degradation. *Genes to Cells* **2**, 13–28.
11. Hicke, L. (1997). Ubiquitin-dependent internalization and down-regulation of plasma membrane proteins. *FASEB J.* **11**, 1215–1226.
12. Hochstrasser, M. (1996). Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**, 405–439.
13. Scheffner, M., Smith, S. & Jentsch, S. (1998) In *Ubiquitin and the Biology of the Cell* (Peters, J.-M., Harris, J.R. & Finley, D., eds), pp. 65–98. Plenum Press, New York.
14. Baumeister, W., Walz, J., Zühl, F. & Seemüller, E. (1998). The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92**, 367–380.
15. Hilt, W. & Wolf, D.H. (1996). Proteasomes: destruction as a programme. *Trends. Biochem. Sci.* **21**, 96–102.
16. Bachmair, A. & Varshavsky, A. (1989). The degradation signal in a short-lived protein. *Cell* **56**, 1019–1032.
17. Chau, V., Tobias, J.W., Bachmair, A., Marriott, D., Ecker, D.J., Gonda, D.K. & Varshavsky, A. (1989). A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* **243**, 1576–1583.
18. Hill, C.P., Johnston, N.L. & Cohen, R.E. (1993). Crystal structure of a ubiquitin-dependent degradation substrate: a three-disulfide form of lysozyme. *Proc. Natl. Acad. Sci. USA* **90**, 4136–4140.
19. Dohmen, R.J., Wu, P. & Varshavsky, A. (1994). Heat-inducible degron: a method for constructing temperature-sensitive mutants. *Science* **263**, 1273–1276.
20. Sikorski, R.S. & Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *S. Cerevisiae*. *Genetics* **122**, 19–27.
21. Johnson, E.S., Bartel, B.W. & Varshavsky, A. (1992). Ubiquitin as a degradation signal. *EMBO J.* **11**, 497–505.
22. Johnson, E.S., Ma, P.C.M., Ota, I.M. & Varshavsky, A. (1995). A proteolytic pathway that recognizes ubiquitin as a degradation signal. *J. Biol. Chem.* **270**, 17442–17456.
23. Lévy, F., Johnsson, N., Rümenapf, T. & Varshavsky, A. (1996). Using ubiquitin to follow the metabolic fate of a protein. *Proc. Natl. Acad. Sci. USA* **93**, 4907–4912.
24. Ghislain, M., Dohmen, R.J., Lévy, F. & Varshavsky, A. (1996). Cdc48p interacts with Ufd3p, a WD-repeat protein required for ubiquitin-dependent proteolysis in *Saccharomyces Cerevisiae*. *EMBO J.* **15**, 4884–4899.
25. Baker, R.T. & Varshavsky, A. (1991). Inhibition of the N-end rule pathway in living cells. *Proc. Natl. Acad. Sci. USA* **87**, 2374–2378.
26. Mosteller, R.D. & Goldstein, B.E. (1984). A mathematical model that applies to protein degradation and posttranslational processing of proteins and to analogous processes for other molecules in non-growing and exponentially growing cells. *J. Theor. Biol.* **108**, 597–621.
27. Johnston, J.A., Johnson, E.S., Waller, P.R.H. & Varshavsky, A. (1995). Methotrexate inhibits proteolysis of dihydrofolate reductase by the N-end rule pathway. *J. Biol. Chem.* **270**, 8172–8178.
28. Frieden, C. (1990). Refolding of Escherichia coli dihydrofolate reductase: sequential formation of substrate-binding sites. *Proc. Natl. Acad. Sci. USA* **87**, 4413–4416.
29. Spence, J., Sadis, S., Haas, A.L. & Finley, D. (1995). A ubiquitin mutant

with specific defects in DNA repair and multiubiquitination. *Mol. Cell. Biol.* **15**, 1265–1273.

30. Gonda, D.K., Bachmair, A., Wünning, I., Tobias, J.W., Lane, W.S. & Varshavsky, A. (1989). Universality and structure of the N-end rule. *J. Biol. Chem.* **264**, 16700–16712.
31. Gu, H., Marth, J.D., Orban, P.C., Mossmann, H. & Rajewski, K. (1994). Deletion of a DNA polymerase β gene segment in T cells using cell type-specific gene targeting. *Science* **265**, 103–106.
32. No, D., Yao, T.-P. & Evans, R.M. (1996). Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. USA* **93**, 3346–3351.
33. St-Onge, L., Furth, P.A. & Gruss, P. (1996). Temporal control of the Cre recombinase in transgenic mice by a tetracycline-responsive promoter. *Nucl. Acids Res.* **24**, 3975–3877.
34. Matthews, D.A., Bolin, J.T., Burridge, J.M., Filman, D.J., Volz, K.W. & Kraut, J. (1985). Dihydrofolate reductase: the stereochemistry of inhibitor selectivity. *J. Biol. Chem.* **260**, 392–399.
35. Oefner, C., D'Arcy, A. & Winkler, F.K. (1988). Crystal structure of human dihydrofolate reductase complexed with folate. *Eur. J. Biochem.* **174**, 377–385.
36. Thillet, J., Absil, J., Stone, S.R. & Pictet, R. (1988). Site-directed mutagenesis of mouse dihydrofolate reductase: mutants with increased resistance to methotrexate and trimethoprim. *J. Biol. Chem.* **263**, 12500–12508.
37. Arkowitz, R.A., Joly, J.C. & Wickner, W. (1992). Translocation can drive the unfolding of a preprotein domain. *EMBO J.* **12**, 243–253.
38. Eilers, M. & Schatz, G. (1986). Binding of a specific ligand inhibits import of a purified precursor into mitochondria. *Nature* **322**, 228–232.
39. Klingenberg, O. & Olsnes, S. (1996). Ability of methotrexate to inhibit translocation to the cytosol of dihydrofolate reductase fused to diphtheria toxin. *Biochem. J.* **313**, 647–653.
40. Madura, K. & Varshavsky, A. (1994). Degradation of Gα by the N-end rule pathway. *Science* **265**, 1454–1458.
41. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A. & Seidman, J.G. & Struhl, K. (1996) *Current Protocols in Molecular Biology*. Wiley-Interscience, New York.
42. Madura, K., Dohmen, R.J. & Varshavsky, A. (1993). N-recognin/Ubc2 interactions in the N-end rule pathway. *J. Biol. Chem.* **268**, 12046–12054.

Cell, Vol. 92, 489–499, February 20, 1998, Copyright ©1998 by Cell Press

Ump1p Is Required for Proper Maturation of the 20S Proteasome and Becomes Its Substrate upon Completion of the Assembly

Paula C. Ramos,* Jörg Höckendorff,*
Erica S. Johnson,†‡ Alexander Varshavsky,†
and R. Jürgen Dohmen*§
*Biotechnologisches Zentrallabor
Institut für Mikrobiologie
Heinrich-Heine-Universität Düsseldorf
Universitätsstr. 1, Geb. 25.02
40225 Düsseldorf
Germany
†Division of Biology
California Institute of Technology
Pasadena, California 91125
‡Laboratory of Cell Biology
The Rockefeller University
New York, New York 10021

Summary

We report the discovery of a short-lived chaperone that is required for the correct maturation of the eukaryotic 20S proteasome and is destroyed at a specific stage of the assembly process. The *S. cerevisiae* Ump1p protein is a component of proteasome precursor complexes containing unprocessed β subunits but is not detected in the mature 20S proteasome. Upon the association of two precursor complexes, Ump1p is encased and is rapidly degraded after the proteolytic sites in the interior of the nascent proteasome are activated. Cells lacking Ump1p exhibit a lack of coordination between the processing of β subunits and proteasome assembly, resulting in functionally impaired proteasomes. We also show that the propeptide of the Pre2p/Doa3p β subunit is required for Ump1p's function in proteasome maturation.

Introduction

Ubiquitin (Ub)–dependent proteolysis underlies the bulk of nonlysosomal protein degradation in eukaryotic cells (reviewed by Hochstrasser, 1996; Varshavsky, 1997). Naturally short-lived as well as damaged or otherwise abnormal proteins are recognized by the Ub system and are marked for degradation by the attachment of multi-Ub chains. Ubiquitylated proteins are degraded by the 26S proteasome, an ~2000 kDa, multisubunit, ATP-dependent protease that consists of the 19S complex, which is required specifically for the degradation of ubiquitylated proteins, and an ~700 kDa complex, called the 20S proteasome, which is the ATP-independent catalytic core of the 26S proteasome (reviewed by Peters, 1994; Coux et al., 1996; Lupas et al. 1997).

The 20S proteasome is universal among eukaryotes; its structural homologs have also been found in archaeons and eubacteria (reviewed by Coux et al., 1996). High-resolution crystal structures have been reported for the 20S proteasomes of the archaeon *Thermoplasma acidophilum* and the eukaryote *Saccharomyces cerevisiae* (Löwe et al., 1995; Groll et al., 1997). The *Thermoplasma* proteasome contains two types of subunits, α and β, which form a hollow cylinder composed of four heptameric rings in the configuration $\alpha_7\beta_7\beta_7\alpha_7$. The yeast 20S proteasome has a similar structure but contains 14 distinct subunits, seven of the α type and seven of the β type. The N-terminal threonines of the β subunits of the *Thermoplasma* proteasome act as nucleophiles in catalyzing the hydrolysis of peptide bonds of polypeptide substrates (Seemüller et al., 1995). Lys-33 and Glu-17 of the β subunit also play a role in catalyzing the cleavage of peptide bonds. The 14 identical β subunits of the *Thermoplasma* proteasome form 14 identical active sites, which catalyze the cleavage of substrates after hydrophobic amino acid residues (chymotrypsin-like active sites). These sites are located on the inner surface of a central chamber formed by the two rings of β subunits (Löwe et al., 1995; Seemüller et al., 1995). β subunits are synthesized as inactive precursors containing a propeptide that is thought to be cleaved off autocatalytically, yielding the mature β subunit bearing N-terminal Thr. The propeptides of β subunits are not required for the in vitro assembly of the *Thermoplasma* proteasome (Seemüller et al., 1996).

In eukaryotic proteasomes, only 3 of the 7 distinct β subunits contain the three conserved residues required for activity of the *Thermoplasma* proteasome. Genetic and structural data suggest that these three subunits provide the active-site nucleophiles for the three distinct catalytic activities of eukaryotic proteasomes, namely the "chymotrypsin-like" (see above), the "trypsin-like" (cleavage after basic residues), and the "peptidylglutamyl peptide hydrolyzing (PGPH)" activity (cleavage after acidic residues). In *S. cerevisiae*, these three β subunits are Pre2p/Doa3p, Pup1p, and Pre3p (Heinemeyer et al., 1993; Chen and Hochstrasser, 1996; Arendt and Hochstrasser, 1997; Groll et al., 1997; Heinemeyer et al., 1997). In the human 20S proteasome, the related proteins are MB1, δ, and Z; during the immune response, these subunits can be replaced by their respective homologs LMP7, LMP2, and MECL-1/LMP10 (reviewed by Coux et al., 1996). In the case of LMP2 and LMP7, it has been demonstrated that these replacements are functionally relevant in altering the specificity of antigen presentation by the MHC class I pathway. The catalytically active β subunits of the eukaryotic proteasome are synthesized with propeptides, similarly to the β subunit in *Thermoplasma* (Schmidtke et al., 1997).

In the crystal structure of the *S. cerevisiae* 20S proteasome, the opening to the proteasome's interior, formed by the outer ring of the α subunits, is not large enough to admit even an unfolded polypeptide chain, let alone a folded protein (Groll et al., 1997). The degradation of larger substrates requires the 19S complexes, which attach at both sides of the 20S proteasome, yielding the 26S proteasome. The 19S complexes contain subunits that bind multi-Ub chains, at least one Ub-specific isopeptidase that disassembles these chains, and several

---

[§] To whom correspondence should be addressed.

ATPases that are thought to be involved in perturbing the substrate conformationally and guiding it to the interior of the 20S proteasome (Deveraux et al., 1994; Hochstrasser et al., 1995; Jentsch and Schlenker, 1995; Coux et al., 1996).

Studies with mammalian cells have shown that the 20S proteasome is assembled through a 15–16S intermediate, apparently a half-proteasome. This intermediate contains all 14 α and β subunits, some of which are in the precursor (propeptide-bearing) form, and several uncharacterized polypeptides as well (Frentzel et al., 1994; Yang et al., 1995; Nandi et al., 1997; Schmidtke et al., 1997). Studies on the yeast 20S proteasome have demonstrated that processing of proPre2p (identical to proDoa3p) is coupled to formation of the 20S proteasome from two half-proteasome precursors (Chen and Hochstrasser, 1996). Formation of the active site capable of autocatalytic processing of proPre2p depends on the juxtaposition of proPre2p and Pre1p on the opposite sides of the two halves of the proteasome. This mechanism is thought to prevent activation of proteolytic sites before the central hydrolytic chamber has been sealed off from the cytosol through association of the two halves of the proteasome (Chen and Hochstrasser, 1996). The Pre2p propeptide is essential for the formation of functional proteasomes. Moreover, this propeptide can operate in *trans*, suggesting that it serves a chaperone-like function in proteasome maturation (Chen and Hochstrasser, 1996).

In the present work, we identify Ump1p, a novel protein, as a component of a precursor complex of the 20S proteasome. This precursor contains unprocessed β subunits. Upon formation of the 20S proteasome from two such precursors, the propeptides of β subunits are removed, a process that is accompanied by removal of Ump1p from the proteasome through Ump1p degradation, which requires the proteasome's proteolytic activity. In *ump1Δ* cells, which lack Ump1p, coordination of 20S complex formation and processing of β subunits is impaired, resulting in incompletely or (in the case of Pre2p) prematurely processed β subunits. These findings reveal Ump1p as a novel type of molecular chaperone, a short-lived maturation factor required for the efficient biogenesis of the 20S proteasome. We also describe genetic evidence that the propeptide of proPre2p is required for Ump1p-dependent proteasome maturation, and we present a model that accounts for some of the functions of Ump1p and propeptides in proteasome maturation.

Results

The *UMP1* Gene Is Required for Ubiquitin-Mediated Proteolysis

To identify the genes required for Ub-dependent proteolysis in *S. cerevisiae*, we screened for mutants defective in the degradation of test substrates. One such mutant, termed *ump1-1* (ub-mediated proteolysis), exhibited defects in the degradation of several normally short-lived proteins (see below). The complementing *UMP1* gene encoded a 148-residue (16.8 kDa) protein. Searches of the databases did not identify close homologs of Ump1p but did detect similarities to regions of several proteins. One intriguing similarity was between Ump1p and C-terminal regions of the protease inhibitor contrapsin (25% identity, 52% similarity) and related proteins (Figure 1A), consistent with the likely role of Ump1p as an inhibitor of premature proteolytic activation of the proteasome precursor complexes (see below).

Pulse-chase analyses in *ump1Δ* mutants revealed a strong stabilization of several normally short-lived test proteins, in particular an N-end rule substrate Arg–β-galactosidase (Arg–β-gal), which bears N-terminal Arg, a destabilizing residue (Varshavsky, 1996), Ub-Pro–β-gal (a substrate of the UFD pathway; Johnson et al., 1995; Varshavsky, 1997), and $\alpha 2_{deg1}$–β-gal (a substrate of the DOA pathway; Hochstrasser and Varshavsky, 1990; Hochstrasser et al., 1995) (Figure 1B). The recognition and ubiquitylation of these substrates involve different recognins (E3 proteins) and different Ub-conjugating (E2) enzymes. The pleiotropic character of the *ump1Δ* phenotype suggested that Ump1p functions downstream of the recognition and ubiquitylation components of the Ub system, perhaps at the step of proteolysis, or in regulating the supply of Ub.

ump1Δ Mutants Are Hypersensitive to a Variety of Stresses, Accumulate Ub-Protein Conjugates, and Do Not Sporulate

ump1Δ mutants grew more slowly than congenic wild-type (wt) cells at 30°C or lower temperatures and were severely growth-impaired at higher temperatures (37°C) (Figure 1C and data not shown). In addition, they were hypersensitive to cadmium ions and to the arginine analog canavanine (Figure 1C). Similar phenotypes have been reported for mutants in genes encoding E2 (Ubc) enzymes or proteasome subunits (Heinemeyer et al., 1993; Jungmann et al., 1993). Comparisons of proteins from whole-cell extracts of wt and *ump1Δ* cells by immunoblotting with anti-Ub antibodies showed a dramatic accumulation of Ub-protein conjugates in the *ump1Δ* mutant. At the same time, the level of free Ub was reduced in *ump1Δ* cells (data not shown). Thus, the primary cause of the *ump1Δ* phenotype appears to be the impaired ability of *ump1Δ* cells to degrade Ub-protein conjugates. Homozygous *ump1Δ* diploids (strain JD61) were unable to sporulate (data not shown).

Ump1p Is a Component of Proteasome Precursors

To investigate Ump1p biochemically, an epitope tag was linked to its C terminus. A single copy of the modified *UMP1* gene (*UMP1-ha*), expressed from its natural promoter and chromosomal location, restored wt growth rates. We analyzed whole-cell extracts of a strain expressing Ump1p-ha and Pre1p-ha (the latter an ha-tagged β subunit of the proteasome) by gel filtration on Superose-6. Ump1p-ha eluted in fractions 22 and 23, both of which also contained Pre1p-ha and other proteasomal subunits (Figures 2A, 2B, and 3A) but lacked the chymotrypsin-like activity of the mature 20S proteasome. The apparent size of the Ump1p-containing complex was 300–400 kDa (Figure 2A). 20S and 26S proteasomes, as well as free 19S caps, and 20S proteasomes with one attached 19S cap, eluted in earlier fractions

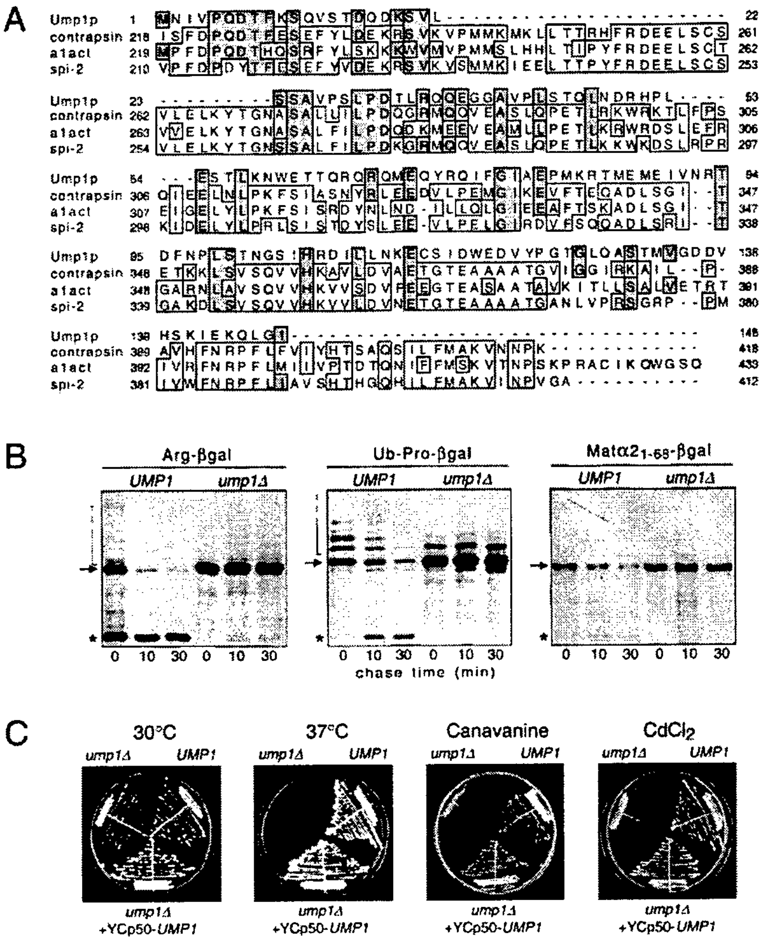


Figure 1. Ump1p Is Required for Ubiquitin-Mediated Proteolysis

(A) Sequence similarities between Ump1p and C-terminal parts of mouse contrapsin, human alpha1 antitrypsin inhibitor (a1act) and mouse spi-2 (PIR accession numbers, respectively, JX0129, A90475, and S31305). The sequences were aligned using the PileUp program (GCG package, version 7.2, Genetics Computer Group, Madison, WI). Gaps (indicated by hyphens) were used to maximize alignments. Residues identical between Ump1p and at least one of the other proteins are shaded in gray. Residues identical among at least two proteins other than Ump1p are boxed.

(B) Pulse-chase analysis comparing the metabolic stabilities of R-β-gal, Ub-P-β-gal, and MATα$2_{1\text{-}68}$-β-gal in wt (*UMP1*) and *ump1Δ* cells. The open-ended brackets denote the position of multiubiquitylated β-gal species. Asterisks denote an ~90 kDa β-gal cleavage product characteristic of short-lived β-gal derivatives (Dohmen et al., 1991).

(C) Growth of *ump1Δ* (JD59), *UMP1* (JD47-13C), and *ump1Δ* cells transformed with YCp50-*UMP1* (expressing wild-type *UMP1*). Cells were streaked on YPD plates and incubated for 2 days at 30°C (or at 37°C where indicated), with 0.8 μg/ml canavanine or 30 μm $CdCl_2$ where indicated.

(15–21), as determined by nondenaturing gel electrophoresis (Figure 2B and data not shown), by assays of the proteasome's proteolytic activity, and by immunological detection of Cim3p, a subunit specific for the 19S cap of the 26S proteasome (Ghislain et al., 1993).

Cochromatography of Ump1-ha and Pre1p-ha in fractions 22 and 23 (Figure 2A) suggested that Ump1p is a component of proteasome precursor complexes. Consistent with this possibility, the Ump1p-containing complex had a higher mobility than the 20S proteasome upon nondenaturing gel electrophoresis. A proteasome precursor with an electrophoretic mobility indistinguishable from that of the Ump1p-containing complex was also detected in extracts from a strain that expressed Pup1p-ha, another β subunit of the 20S proteasome (Figure 2B).

To produce independent evidence bearing on the nature of the Ump1p-containing complex, we constructed a strain that expressed both Ump1p-ha and doubly tagged Pre1p-Flag-His_6 (Pre1p-FH). Affinity chromatography on Ni-NTA-agarose and anti-Flag antibody resin was then used to purify complexes containing Pre1p-FH. Ump1p-ha cofractionated with Pre1p-FH in both affinity purification steps (Figures 2C and 2D), confirming that Ump1p is a component of a distinct proteasome-related complex. This complex sedimented at ~15S in sucrose gradients (data not shown). Taken together, this evidence strongly suggested that the Ump1p-containing complex is a precursor of the 20S proteasome that is similar to the 15–16S precursors observed in the maturation pathway of mammalian proteasomes (Frentzel et al., 1994; Yang et al., 1995).

The Ump1p Proteasome Precursor Complex Contains Unprocessed β Subunits

Several subunits of the proteasome are synthesized as precursors containing N-terminal extensions (propeptides) that are absent from the mature proteasome. These precursors are detected in the mammalian 15–16S complex (Frentzel et al., 1994; Yang et al., 1995). To compare the Ump1p-containing complex with the mammalian 15–16S complex, we analyzed the former for the presence of propeptide-containing proteasomal subunits. These β subunits were modified by the addition of C-terminal ha tags. Strains expressing Pup1p-ha, Pre2p-ha, and Pre3p-ha grew at wt rates (data not shown). When extracts of these strains were fractionated by gel filtration on Superose-6, the fractions containing the Ump1p complex contained largely the β subunit precursors, proPup1-ha, proPre2p-ha, and proPre3p-ha (Figure 3A). The corresponding mature β subunits largely eluted in the earlier fractions that contained 20S and 26S proteasomes (Figure 3A). Subtle but reproducible differences

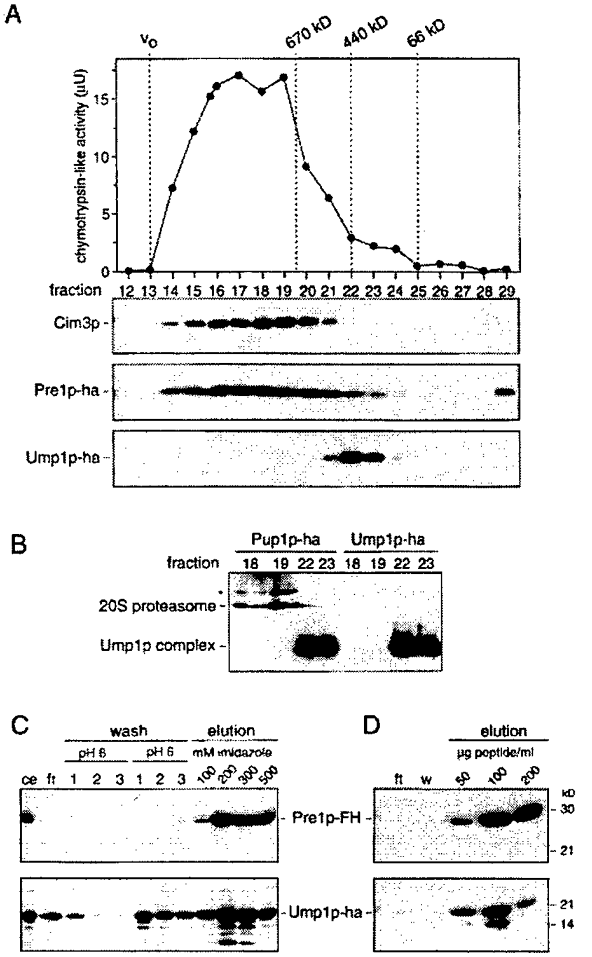


Figure 2. Ump1p Is a Component of Proteasome Precursor Complexes

(A) An extract from strain JD127 expressing Ump1p-ha and Pre1p-ha instead of wt versions of these proteins was fractionated by gel filtration on Superose-6. The upper panel shows the results of measurements of chymotrypsin-like activity in the relevant fractions (12–29). The last fraction, representing the void volume (v_o), and positions of the peaks of marker proteins are indicated by dashed lines. The same fractions were analyzed by SDS-PAGE and immunoblotting with anti-Cim3p and anti-ha antibodies, which detected Cim3p, Pre1p-ha, and Ump1p-ha, as indicated.

(B) Analysis of Superose-fractionated 20S proteasomes and precursor complexes by nondenaturing gel electrophoresis and immunoblotting with anti-ha antibody. The strains were JD139 and JD129 expressing, respectively, Pup1p-ha and Ump1p-ha. Asterisk, 20S proteasome with one attached 19S regulator cap. The 26S proteasome eluted in fractions 14–17 of Superose-6 (not shown).

(C) Affinity purification of Pre1p-FH on Ni-(NTA)-Sepharose and copurification of Ump1p-ha. Purification of proteins from an extract of strain JD126 expressing Ump1p-ha and Pre1p-FH was carried out as described in the Experimental Procedures. Different fractions were analyzed by SDS-PAGE and immunoblotting with anti-Flag (upper panel) and anti-ha antibodies (lower panel). Ce, crude extract; ft, flow-through. In contrast, no binding of Ump1p to the resin was observed when extracts from strain JD129 expressing untagged Pre1p were used (not shown).

(D) Affinity purification of Pre1p-FH on anti-Flag antibody agarose resin. Material that was eluted from the Ni-(NTA)-Sepharose column shown in panel C was subjected to a second affinity purification on anti-Flag antibody resin. Pre1p-FH was specifically eluted with increasing concentrations of the Flag peptide. W, wash.

were detected among the tagged β subunits with respect to the distribution of their precursor and processed forms in various fractions. For example, pro-Pup1p-ha was detectable only in fractions 22 and 23, and was absent from fraction 21. By contrast, proPre2p-ha was detectable in the larger complexes (down to fraction 19), while fractions 22 and 23 already contained some processed Pre3p-ha. These patterns suggested a defined order of processing events. One conclusion from these experiments is that the Ump1p-containing complex is a precursor of the 20S proteasome that contains unprocessed β subunits and is therefore proteolytically inactive (Figures 3A and 3C).

Ump1p Is Required for Correct Proteasome Maturation

Next, we asked what effect the *ump1Δ* mutation has on maturation and activity of the proteasome. Specifically, the analyses described above were repeated with extracts obtained from *ump1Δ* cells. Figure 3C shows that there was a significant reduction in the three proteolytic activities of the proteasome in the fractions containing the 20S and 26S proteasomes (fractions 14–21). The reduction of the specific activity appears to be partially compensated by an increased expression of proteasomes (Figures 3A and 3B, data not shown) similar to that observed in cells expressing mutant β subunits (Arendt and Hochstrasser, 1997). In addition, a new (absent from wt cells) peak of chymotrypsin-like activity encompassing fractions 22 and 23 was detected in extracts from *ump1Δ* cells. These were the fractions that contained the Ump1p complex from wt extracts (see above and Figure 3A), suggesting that the absence of Ump1p from the proteasome precursor complex was the cause for premature activation of its chymotryptic activity in the *ump1Δ* mutant. To verify that the detected activity was actually of the ~15S precursor complex from *ump1Δ* cells, the fraction 22 samples from strains expressing either wt Pre2p or Pre2p-ha were incubated with anti-ha antibody and protein A Sepharose. This treatment did not deplete the chymotryptic activity from the sample containing untagged Pre2p but did deplete ~80% of the activity from an otherwise identical sample derived from *ump1Δ* cells expressing Pre2p-ha (data not shown). Thus, the chymotryptic activity of this fraction resided in the proteasome precursor complexes.

To follow the appearance of the precursor and mature forms of β subunits in proteasome precursors and mature proteasomes, we analyzed extracts from *ump1Δ* cells expressing ha-tagged versions of these subunits by immunoblotting with anti-ha antibody. The processing of the three analyzed β subunits, Pup1p-ha, Pre2p-ha, and Pre3p-ha, was strikingly different in *ump1Δ* and wt cells (Figure 3A). For all three subunits, a dramatic increase of their precursors was detected in the fractions (14–21) that contained the 20S and 26S forms of the proteasome. In addition, different processed variants, possibly representing processing intermediates of pro-Pre2p, could be detected in fractions 22 and 23, which contained the ~15S proteasome precursor, and in some of the proteasome-containing fractions as well (Figure 3A). In contrast, with the exception of Pre3p, almost no processed β subunits were present in the fractions

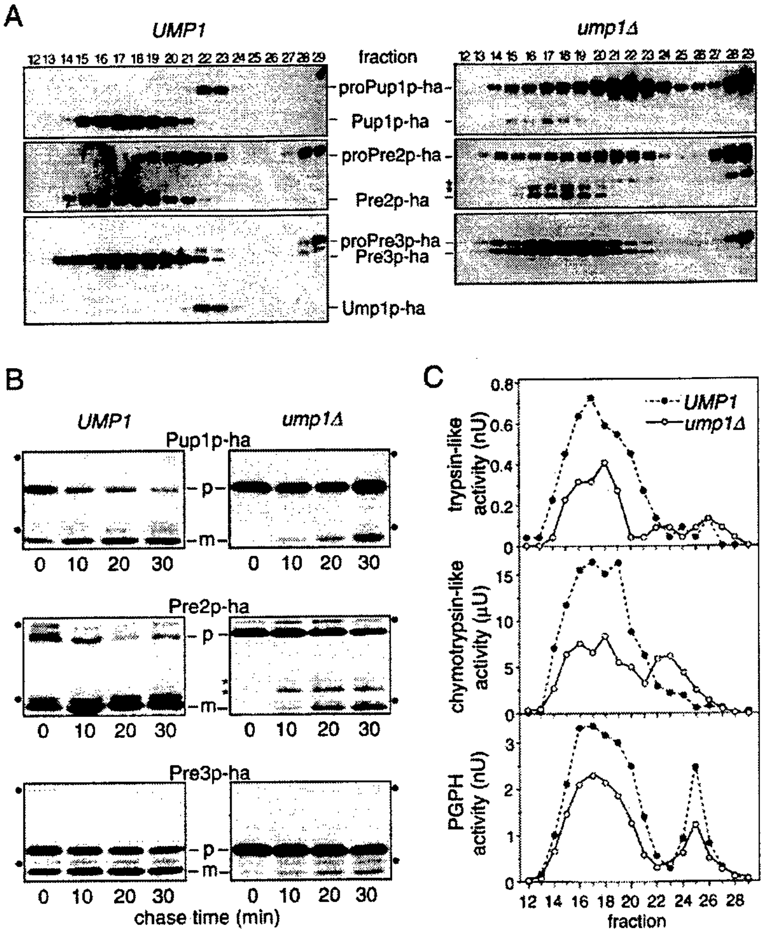


Figure 3. Ump1p Is Required for the Correct Processing of Proteasomal β Subunits

(A) SDS-PAGE and anti-ha immunoblot analyses of extracts fractionated on Superose-6. The congenic yeast strains used were JD131, JD132, JD133, JD134, JD135, and JD136 (Table 1). Western blots of the *ump1Δ* extracts were developed for a shorter time to compensate for an increased expression of proteasomal subunits in the mutant. Asterisks, processing intermediates or incompletely processed forms of Pre2p-ha that were observed in the *ump1Δ* mutant but not in wt cells.

(B) Pulse-chase analysis comparing the rate of β subunit processing in wt cells to that in *ump1Δ* cells (same strains as in A). Cells were pulse-labeled with ^{35}S-Met/Cys for 5 min. Proteins were precipitated with anti-ha antibody. Fluorographic exposures were $\sim$3 times longer for the *UMP1* samples. P, propeptide-containing precursor form; m, mature form; asterisks, as in (A); dots, nonspecific bands present in some samples.

(C) Comparison of proteolytic activities in extracts with equal amounts of total protein of *UMP1* (JD127) and *ump1Δ* cells (JD75) fractionated by gel filtration on Superose-6. The profile of the chymotrypsin-like activity in *UMP1* is the same as in Figure 2A.

containing the $\sim$15S proteasome precursor from wt cells, and no processing intermediates of Pre2p could be detected either (Figure 3A). These findings indicate that Ump1p is required to prevent premature processing of at least proPre2p and that the presence of Ump1p is important for the coordination of proteasome assembly and subunit processing.

A defect of *ump1Δ* cells in the processing of β subunits was also observed by following the metabolic fate of these subunits in pulse-chase experiments. In wt cells, most of the propeptide-containing forms of β subunits (proPup1-ha, proPre2p-ha and proPre3p-ha) were converted into their mature counterparts during the 30-min chase (Figure 3B). By contrast, in the *ump1Δ* mutant, the bulk of the β subunit precursors remained unprocessed during this time, a finding consistent with the results of immunoblot analyses (Figure 3A). In the case of Pre2p-ha, pulse-chase analysis again revealed the species of intermediate size (putative processing intermediates) in extracts from *ump1Δ* cells. These species were absent from the equivalent samples derived from wt cells (Figure 3B).

Ump1p Is Degraded upon Formation of the 20S Proteasome

The experiments above demonstrated that Ump1p is a component of proteasome precursors that is absent from the mature 20S proteasome. One possibility was that Ump1p leaves the precursor complex upon formation of the 20S proteasome from the two $\sim$15S precursors and may then assist with another round of proteasome assembly, thus acting catalytically. It was also possible that formation of the 20S proteasome sterically traps Ump1p. Since the formation of the 20S form of the proteasome coincides with the appearance of its proteolytic activities (Frentzel et al., 1994; Chen and Hochstrasser, 1996), Ump1p might be degraded by the newly formed 20S proteasome. Pulse-chase analysis of Ump1p-ha in wt cells showed that Ump1p-ha is indeed degraded in vivo, the rate of its degradation being similar to the rate of disappearance of proPup1p-ha, which is converted to Pup1p-ha upon maturation of the proteasome (compare Figures 4A and 3B).

Ump1p Is Stabilized in Proteasome Mutants and Can Be Detected Inside the 20S Particle

If Ump1p becomes a substrate of the newly formed 20S proteasome, the degradation of Ump1p should be inhibited by mutations that affect the proteasome's proteolytic activities. We used pulse-chase assays to follow the metabolic fate of Ump1p-ha in the *pre1-1* mutant, which is known to be deficient in the chymotryptic activity of the proteasome (Heinemeyer et al., 1993). Ump1p was partially stabilized in *pre1-1* cells, as indicated by the increased amount of ^{35}S-labeled Ump1p in *pre1-1* extracts versus wt extracts that contained the same total amount of TCA-precipitable ^{35}S (Figure 4A). However, we still observed a significant decrease of pulse-labeled Ump1p upon increasing chase times when the immunoprecipitation was carried out under nondenaturing conditions (Figure 4A). By contrast, when the extracts were treated with 0.4% SDS at 100°C before immunoprecipitation, virtually no decrease of the Ump1p signal during the chase was observed in *pre1-1* cells. These results were consistent with the possibility that the newly formed Ump1p-ha became inaccessible to the anti-ha antibody during the chase because it became trapped within the newly formed *pre1-1* proteasome. If so, Ump1p was expected to be present in fractions from the Superose-6 column that corresponded to the 20S and 26S proteasomes in the *pre1-1* mutant. Indeed, whereas in wt (*PRE1*) cells Ump1p-ha was detected by SDS-PAGE and immunoblotting only in fractions 22 and 23 (corresponding to the ~15S proteasome precursor complex), in the mutant (*pre1-1*) cells Ump1p-ha was detected in fractions 15–23, indicating that it was also present in mature proteasomes (Figure 4B). We assayed the accessibility of Ump1p-ha to anti-ha antibody by immunoprecipitation from fraction 19 (the 20S proteasome) of extracts from the *pre1-1* mutant and from fraction 22 (the ~15S proteasome precursor) of extract from *PRE1* cells. Ump1p-ha could be immunoprecipitated from either fraction 19 or 22 following pretreatment with 0.1% SDS, but not under nondenaturing conditions (Figure 4C), indicating that the ha tag was inaccessible to the anti-ha antibody in both the *pre1-1* proteasome and the precursor complex. However, when otherwise identical assays were carried out with a polyclonal anti-Ump1p antiserum, most of Ump1p could be immunoprecipitated under nondenaturing conditions from fraction 22 (the ~15S proteasome precursor) but was still not immunoprecipitated from fraction 19 (the 20S *pre1-1* proteasome). In contrast, when polyclonal anti-proteasome antibodies were used, Ump1p was precipitated quantitatively along with both complexes (Figure 4C).

These results indicated that a part of Ump1p other than its C terminus is accessible to anti-Ump1p antibodies in the ~15S precursor. However, Ump1p becomes entirely inaccessible (under nondenaturing conditions) upon formation of the 20S proteasome from two ~15S precursors. This interpretation was supported by examining the sensitivity of Ump1p-ha in different complexes to trypsin digestion (Figure 4D). Specifically, Ump1p-ha in the 20S proteasomes (fraction 19) from *pre1-1* cells was completely protected against trypsin. In contrast, Ump1p-ha in the ~15S proteasome precursors was detectably accessible to trypsin. The degradation of Ump1p was incomplete, resulting in a protected fragment of Ump1p-ha that lacked the ~5 kDa N-terminal region but retained the C-terminal ha tag (Figure 4D). Interestingly, in a similar experiment with a strain expressing pro-Pre2p-ha, we observed that, in the Ump1-containing ~15S proteasome precursor complex (fraction 22), pro-Pre2p-ha was shortened by trypsin treatment to yield a product whose electrophoretic mobility was indistinguishable from that of the natural mature Pre2p-ha (Figure 4D). In these experiments, the overall structure of the Ump1p proteasome precursor complex remained intact as judged by native gel analysis of trypsin-treated material (data not shown).

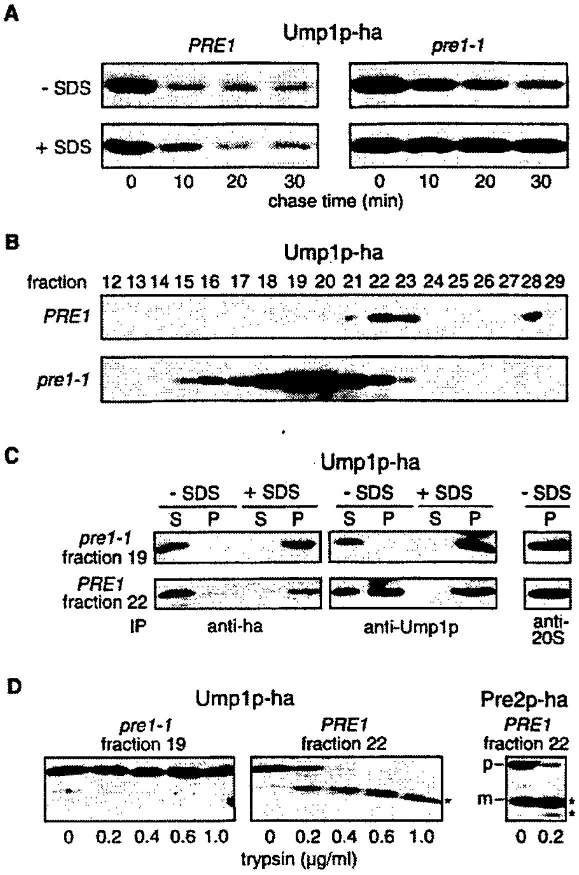


Figure 4. Ump1p Is Stabilized in *pre1-1* Mutants Cells that Are Defective in the Proteasome's Chymotrypsin-like Activity and Persists in 20S Proteasomes

(A) Pulse-chase analysis of Ump1p-ha in wt and *pre1-1* cells. The strains used were JD150 (*PRE1*) and JD151 (*pre1-1*) (Table 1), both expressing Ump1p-ha. The cells were pulse-labeled with ^{35}S-Met/Cys for 5 min. Extracts were prepared from samples taken at different chase times. Samples were then split in halves. One-half (−SDS) was subjected to immunoprecipitation with anti-ha antibody according to the standard pulse-chase protocol. The other half (+SDS) was adjusted to 0.4% SDS and incubated at 100°C for 5 min, then diluted with extraction buffer to the final concentration of 0.1% SDS prior to immunoprecipitation.

(B–D) SDS-PAGE and anti-ha immunoblot analyses. (B) Detection of Ump1p-ha-containing complexes in extracts fractionated by Superose-6 gel filtration. Strains were the same as in (A). Note the accumulation of Ump1p-ha in fractions containing the 20S and 26S proteasome (fractions 15–21) in *pre1-1*. (C) Immunoblot analyses of immunoprecipitations were carried out with material from Superose-6 fraction 19 of extracts from strain JD151 and fraction 22 of extracts from strain JD150, using the indicated antibodies. Immunoprecipitations were performed without addition of SDS (−SDS) or after boiling in the presence of 0.1% SDS (+SDS). P, precipitates; S, supernatants. (D) Assaying trypsin sensitivity of the same material as in (C), and of the Superose-6 fraction 22, from strain JD138 expressing Pre2p-ha. P, proPre2p-ha; m, mature Pre2p-ha; asterisks mark trypsin digestion products.

In the *ump1Δ* Mutant, the Propeptide of Pre2p Is Not Required for Incorporation of Pre2p into the Proteasome

Chen and Hochstrasser (1996) have elegantly demonstrated that the propeptide of Pre2p, if separated from the mature Pre2p, could function in *trans* and thereby

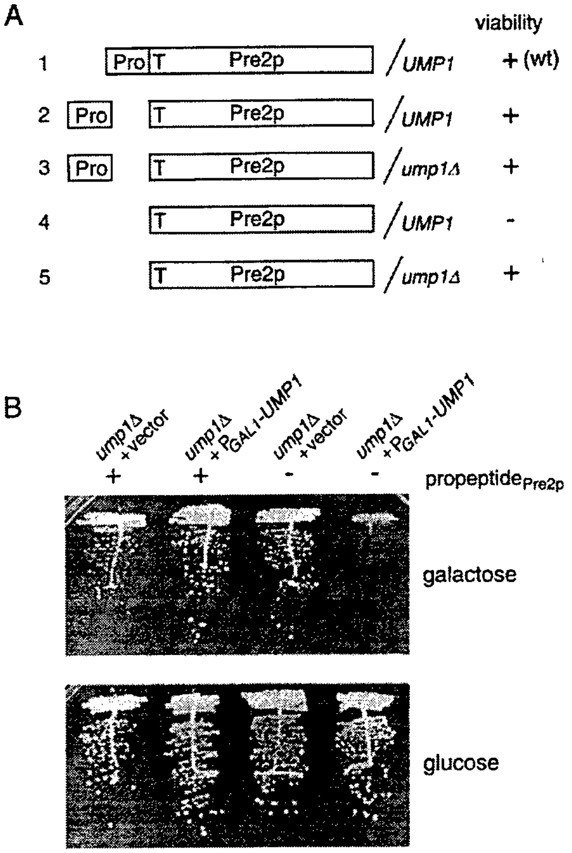


Figure 5. The Pre2p Propeptide Is Not Essential in the *ump*Δ Mutant

(A) Representation of different genotypes of a set of congenic strains and their effects on cell viability. Strain 1 (JD50-10B) is the wt control. Strain 2 (MHY952) expresses Pre2p/Doa3p-ΔLS with N-terminal Thr (T) as a fusion to ubiquitin (not shown), and the Pre2p/Doa3p propeptide (LS) from two separate plasmids (Chen and Hochstrasser, 1996). Strain 3 (JD160), same as strain 2 but *ump1*Δ. Strain 4 would be derived from strain 2 through the loss of the plasmid encoding Pre2p/Doa3p-LS that operates in *trans*. We were unable to produce isolates that lost this plasmid, confirming the earlier demonstration that the propeptide is essential for cell viability (Chen and Hochstrasser, 1996). Strain 5 (JD163) is derived from strain 3 through the loss of the propeptide-encoding plasmid. Loss of the plasmid did not affect cell viability in the *ump1*Δ background.

(B) Induction of *UMP1* expression in the *ump1*Δ background inhibits the growth of a strain that lacks the Pre2p propeptide. Strains 3 (plus) and 5 (minus) shown in (A) were transformed with pJD429 that expressed *UMP1* from the galactose-inducible, glucose-repressible P_{GAL1} promoter, or with an empty vector as a control. Transformants selected on glucose media were pregrown on selective media containing raffinose as a carbon source, then streaked onto selective media with 2% galactose or 2% glucose, and grown for 3 days at 30°C.

still allow the incorporation of Pre2p into the 20S proteasome. Under these conditions, expression of the propeptide is essential for cell viability, suggesting that the propeptide, in *cis* or at least in *trans*, is required for the assembly of an active proteasome (Chen and Hochstrasser, 1996).

We confirmed this result using their strain MHY952, in which mature Pre2p and the propeptide are expressed on separate plasmids (Figure 5A). Specifically, under nonselective growth conditions this strain did not yield viable cells that had lost the plasmid expressing the propeptide. Surprisingly, however, when we constructed and examined a congenic *ump1*Δ derivative of this strain, we noticed that it lost the propeptide-expressing plasmid at a frequency suggesting that this plasmid did not provide a significant growth advantage to cells (data not shown). Indeed, growth rates of cells with the plasmid were indistinguishable from those lacking it (Figure 5B). This result demonstrated that *ump1*Δ is a suppressor of a deletion of the propeptide of Pre2p. If so, expression of *UMP1* in the (*ump1*Δ *PRE2-*Δ*pro*) background should be incompatible with cell viability. This prediction was confirmed when we examined growth properties of the (*ump1*Δ *PRE2-*Δ*pro*) strain transformed with a plasmid expressing *UMP1* from the galactose-inducible P_{GAL1} promoter. On glucose-containing media, the growth rate of this transformant was indistinguishable from that of an otherwise identical transformant carrying the propeptide-expressing plasmid. In contrast, on galactose-containing media, the (*ump1*Δ *PRE2-*Δ*pro*) mutant containing the P_{GAL1}-*UMP1* plasmid was unable to grow, whereas the control strains grew (Figure 5B). This result indicated that the propeptide of Pre2p is essential for the Ump1p-assisted maturation of the proteasome but is not essential for (partially) defective maturation of the proteasome that takes place in the absence of Ump1p.

Discussion

Ump1p, a Novel Maturation Factor of the 20S Proteasome

We describe the discovery of a proteasome maturation factor, termed Ump1p, whose unique properties include a short in vivo half-life that is due to its degradation within the newly formed proteasome. Ump1p is required for coordination of the proteasome's physical assembly and enzymatic activation. In addition, the normally essential propeptide of the Pre2p β subunit was found to become nonessential in the absence of Ump1p. We report the following specific results.

ump1Δ *Mutants Are Defective in Ub-Mediated Proteolysis*

They are sensitive to a variety of stresses and accumulate Ub-protein conjugates (Figure 1 and data not shown). All of the observed phenotypes of *ump1*Δ cells are consistent with the conclusion that *ump1* mutants are impaired in proteasome biogenesis and, consequently, in the degradation of ubiquitylated proteins.

The* UMP1 *Gene

It encodes a polypeptide with a calculated molecular mass of 16.8 kDa. No close sequence homologs of Ump1p were detected in the current databases. However, the presence of a small protein similar in size to Ump1p in preparations of the half-proteasome precursors in mammals (Frentzel et al., 1994; Yang et al., 1995; Nandi et al., 1997; Schmidtke et al., 1997) suggests the presence of a functional homolog of Ump1p in the mammalian proteasome maturation pathway.

Ump1p Is a Component of Proteasome Precursors that Contain Unprocessed* β *Subunits

The Ump1p-containing complex has a molecular mass of 300–400 kDa, sediments at ~15S, migrates significantly faster in native gels than the 20S proteasome,

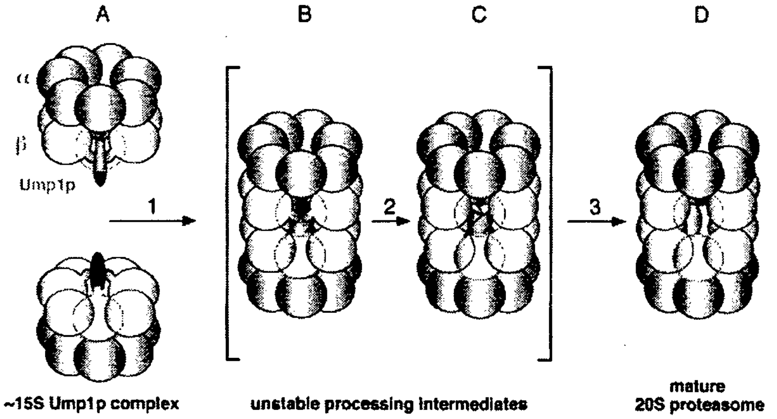


Figure 6. A Model of the Ump1p Function in Proteasome Maturation

Shown is a schematic view of the 20S proteasome and its precursor forms, with the α and β subunits as blue and green balls, respectively. Three of the β subunits are drawn with extensions that represent propeptides. The β subunits in the front are drawn transparent in order to allow a view into the interior chamber of the proteasome. Structure A is a proteasome precursor complex (half proteasome), characterized by the presence of Ump1p and unprocessed β subunits. In step 1, two of these precursors join to build structure B, a step that leads to conformational or positional shifts of Ump1p and propeptides. Conformational changes of the propeptides trigger their autocatalytic processing (step 2), and activation of the proteasome's proteolytic activities. This leads to the degradation of the chamber-entrapped Ump1p by the newly formed proteasome (step 3). Only structures A and D are long-lived enough to be detected in wt cells. The findings with the *pre1-1* mutant, in which the degradation of Ump1p is inhibited, suggest the existence of the short-lived intermediates B and C. See text for details.

and has a subunit composition that is highly similar to that of the 20S proteasome, as determined by SDS-PAGE and immunoblot analyses with anti-20S proteasome antibodies (Figure 2 and data not shown). Taken together, these results indicate that the Ump1p is contained within a half-proteasome precursor complex, whose counterpart has been described in the maturation pathway of mammalian proteasomes (Frentzel et al., 1994; Yang et al., 1995; Nandi et al., 1997; Schmidtke et al., 1997).

Maturation of Pup1p, Pre2p, and Pre3p Is Strongly Impaired in the ump1Δ *Mutant*

Considerably increased amounts of the unprocessed (propeptide-containing) forms of these β subunits were detected in the 20S and 26S fractions from *ump1Δ* cells (Figure 3). In addition, the proteasomes assembled in *ump1Δ* cells had reduced activity. Thus, Ump1p is required for the correct and efficient maturation of the proteasomes.

Partially Processed Forms of Pro-Pre2p Are Detected in the Proteasome Precursor Complex and the Proteasome of ump1Δ *Cells but Not in WT Cells*

These incompletely and prematurely processed forms of Pre2p (Figures 3A and 3B) appear to underlie a chymotrypsin-like activity that was associated with the precursor complex in *ump1Δ* cells but was absent from the same complex of wt cells (Figure 3C). This was suggested by the observation that this activity was not inhibited by treatment with lactacystin and was absent from cells lacking the Pre2p propeptide (P. C. R. and R. J. D., unpublished data). These findings indicated that Ump1p has a dual role in proteasome maturation. Specifically, Ump1p prevents premature processing of proPre2p in the precursor complex and is also required for the correct maturation of active sites upon assembly of the proteasome (see below).

Ump1p Is Degraded by the Newly Formed Proteasome

The kinetics of the rapid degradation of Ump1p is similar to the kinetics of processing of Pup1p (Figures 3B and 4A), suggesting that Ump1p is destroyed by the proteasome upon its formation from the two half-proteasome precursors and the accompanying maturation of active sites (Chen and Hochstrasser, 1996). This conclusion was strongly supported by the observation that Ump1p is significantly stabilized in the *pre1-1* mutant, which is deficient in the proteasome's chymotrypsin-like activity (Figure 4A). Specifically, in contrast to the pattern in wt cells, Ump1p was detectable in the 20S and 26S forms of the *pre1-1* proteasome. Ump1p in these complexes was shielded both from detection by antibodies and from digestion by trypsin (Figures 4B–4D). These and related data suggested a model in which Ump1p is encased within the 20S proteasome upon its assembly from the two Ump1p-containing half-proteasome precursor complexes (Figure 6). Formation of the 20S structure triggers active site maturation, resulting in the degradation of Ump1p.

The Propeptide of Pre2p Is Not Essential in ump1Δ *Cells*

Chen and Hochstrasser (1996) demonstrated that the propeptide of Pre2p is essential for cell viability and that this propeptide can operate in *trans*. They concluded that the propeptide, in addition to rendering proteasome precursors proteolytically inactive, serves a chaperone-like function required for the correct incorporation of Pre2p into the proteasome. One striking result of the present work is that the propeptide of Pre2p becomes dispensable in the *ump1Δ* mutant (Figure 5). In fact, the presence of the propeptide, which is essential for viability of wt cells, does not provide a growth advantage to *ump1Δ* cells that express it.

On the Functions of Ump1p and β Subunit Propeptides in Proteasome Maturation

Our results suggested a model illustrated in Figure 6. This model is based in part on the idea described by Chen and Hochstrasser (1996)—that the active sites of the proteasome are formed upon its assembly from two half-proteasome precursors, through a juxtaposition of subunits at the interface of the dyad-related halves. The advantage of this mechanism is that autocatalytic maturation of the active sites (through processing of the relevant β subunits) is coupled to the assembly of the proteasome, thus avoiding premature processing of the propeptides of β subunits. This view was supported by

Table 1. Yeast Strains

Strain	Relevant Genotype	Source/Comment
YPH500	*MATα ade2-101 his3-Δ200 leu2-Δ1 lys2-801 trp1-Δ63 ura3-52*	Sikorski and Hieter, 1989
JDC9-11	*MATα ump1-1*	Derivative of YPH500
JD47-13C	*MATa his3-Δ200 leu2-3,112 lys2-801 trp1-Δ63 ura3-52*	Dohmen et al., 1995
JD59	*MATa ump1-Δ1::HIS3*	Derivative of JD47-13C
JD53	*MATα*	Derivative of JD47-13C
JD81-1A	*MATα ump1-Δ1::HIS3*	Derivative of JD53
JD75	*MATa PRE1-ha::YIplac211 ump1-Δ1::HIS3*	Derivative of JD47-13C
JD126	*MATa UMP1-ha::YIplac128 PRE1-Flag-6His::YIplac211*	Derivative of JD47-13C
JD127	*MATa UMP1-ha::YIplac128 PRE1-ha::YIplac211*	Derivative of JD71
JD129	*MATa UMP1-ha::YIplac128*	Derivative of JD127
JD131	*MATa UMP1-ha::YIplac128 PUP1-ha::YIplac211*	Derivative of JD129
JD132	*MATa ump1-Δ1::HIS3 PUP1-ha::YIplac211*	Derivative of JD59
JD133	*MATa UMP1-ha::YIplac128 PRE2-ha::YIplac211*	Derivative of JD129
JD134	*MATa ump1-Δ1::HIS3 PRE2-ha::YIplac211*	Derivative of JD59
JD135	*MATa UMP1-ha::YIplac128 PRE3-ha::YIplac211*	Derivative of JD129
JD136	*MATa ump1-Δ1::HIS3 PRE3-ha::YIplac211*	Derivative of JD59
JD138	*MATa PRE2-ha::YIplac211*	Derivative of JD47-13C
JD139	*MATa PUP1-ha::YIplac211*	Derivative of JD47-13C
BBY45	*MATa his3-Δ200 leu2-3,112 lys2-801 trp1-1 ura3-52*	Bartel et al., 1990
JD61	*MATa/MATα ump1-Δ1::HIS3/ump1-Δ2::LEU2*	Congenic with BBY45
JD50-10B	*MATα leu2-3,112 his3-Δ200 trp1-1 ura3-52*	Congenic with BBY45
MHY952	*MATα leu2-3,112 his3-Δ200 trp1-1 ura3-52 pre2(=doa3)Δ1::HIS3* (YCpUbDOA3ΔLS-His) (YEpDOA3$_{LS}$)	Chen and Hochstrasser, 1996, congenic with BBY45
JD160	*MATα ump1-Δ3 pre2(=doa3)Δ1::HIS3* (YCpUbDOA3ΔLS-His) (YEpDOA3$_{LS}$)	Derivative of MHY952
JD163	*MATα ump1-Δ3 pre2(=doa3)-Δ1::HIS3* (YCpUbDOA3ΔLS-His)	Derivative of JD160, cured of YEpDOA3$_{LS}$
WCG4a	*MATα ura3 his3-11 leu2-3,112*	Heinemeyer et al., 1993
YHI29/1	*MATα pre1-1 ura3 his3-11 leu2-3,112*	Heinemeyer et al., 1993
JD150	*MATα UMP1-ha::YIplac128*	Derivative of WCG4a
JD151	*MATα pre1-1 UMP1-ha::YIplac128*	Derivative of YH129/1

biochemical analyses of the proteasome maturation in mammalian cells (Frentzel et al., 1994; Yang et al., 1995; Nandi et al., 1997; Schmidtke et al., 1997).

The results of the present work further support this model. In addition, we discovered that a novel factor, Ump1p, is required for autocatalytic active site maturation and proteasome assembly to occur in a coordinated fashion. A model that accounts for our findings and is consistent with the available evidence is illustrated in Figure 6. In this model, Ump1p has a chaperone-like function required for the efficient processing of the propeptides of β subunits upon the assembly of 20S proteasomes. Specifically, Ump1p interacts with these β subunits, most likely with their propeptides (structure A in Figure 6). Upon formation of the 20S proteasome (Figure 6, step 1), a conformational change of Ump1p, or a change of its position within the complex, would induce a conformational change of the propeptide that results in autocatalytic processing (Figure 6, step 2).

Why is the propeptide of Pre2p essential for proteasome assembly and activation *only* in the presence of Ump1p? We propose that the propeptide of Pre2p is required to induce an alteration of conformation or position of Ump1p upon the assembly of the 20S particle. In this model, the absence of the propeptide would leave Ump1p in a position that is incompatible with the formation of active proteasome. It is possible (and remains to be verified) that this chaperone-like function of a propeptide is unique to Pre2p, as it contains a much longer propeptide (75 residues) than the other β subunits.

In the model of Figure 6, Ump1p is a metabolically unstable chaperone that is required for the proper (and properly timed) processing of β subunits upon the assembly of the 20S proteasome and that is destroyed within the newly activated proteasome. The unusual mechanics of Ump1p, its noncatalytic mode of action, and its degradation by the protease it helps to activate characterize Ump1p as a novel type of molecular chaperone.

Experimental Procedures

Yeast Media

Yeast rich (YPD) and synthetic (S) minimal media with 2% dextrose (SD) or 2% galactose (SG) were prepared as described (Dohmen et al., 1995).

Isolation of the *UMP1* Gene

UMP1 was identified using a selection-based screen for mutants in the N-end rule pathway (reviewed by Varshavsky, 1996). *S. cerevisiae* strain YPH500 (Table 1) was transformed with two plasmids: pJD205, which expressed a Ura3p-based N-end rule reporter substrate (Arg-Tpi1p-Ura3p), and pRL2, which expressed Arg-β-gal, another N-end rule reporter substrate that can be monitored using X-Gal plate assays (Baker and Varshavsky, 1995). Previous work (Dohmen et al., 1994) has shown that rapid degradation of a Ura3-based N-end rule substrate renders the cells Ura$^-$, whereas mutants in the N-end rule pathway that express the same Ura3p-based reporter are Ura$^+$. In this screen, Ura$^+$ isolates were selected on SD media lacking uracil and were then tested on X-Gal plates to verify that the same isolates were also defective in the degradation of the Arg-β-gal N-end rule substrate. One of the mutants thus identified (*ump1-1*) was cured of pJD205 and transformed with a genomic yeast library (Rose et al., 1987). Six transformants yielded plasmids with four overlapping inserts that could restore the ability of cells to degrade Arg-β-gal. A 1981 bp BamHI/Sau3A fragment common to all of the inserts was sequenced and found to contain one complete ORF and two flanking incomplete ORFs. Further mapping, using subcloning and PCR, confirmed that the complete ORF of 445 bp was responsible for the complementation. The *UMP1* sequence (EMBL database accession number: AJ002557) is identical to ORF YBR173C, an ORF subsequently identified by the yeast genome project.

Construction of Yeast Strains and Plasmids

Table 1 lists the strains used in this study. To construct *ump1Δ* alleles, the 1981 bp fragment described above was subcloned into M13mp19. Using single-stranded DNA of the resulting phage, a synthetic oligonucleotide and T4 DNA polymerase, the *UMP1* ORF was precisely deleted and replaced by a BglII restriction site. The resulting fragment was subcloned into pUC19, and the BglII site was used to insert fragments containing the *HIS3, LEU2*, or the *URA3* gene, the latter one being flanked by two direct repeats of a segment of the *E. coli hisG* gene (Alani et al., 1987). The resulting deletion alleles (*ump1-Δ1::HIS3, ump1-Δ2::LEU2*, and *ump1-Δ3::URA3*) were isolated as BamHI fragments, introduced into *S. cerevisiae*, and used to delete the *UMP1* gene (Rothstein, 1991). *ump1Δ* strains carrying an unmarked *ump1-Δ3* allele (resulting from recombination between the *hisG* repeats) were selected on plates containing 5-fluoroorotic acid (Alani et al., 1987). Construction of chromosomal ORFs that expressed C terminally tagged versions of Ump1p or proteasome subunits (Pre1p, Pre2p, Pre3p, and Pup1p) instead of their wt counterparts was performed as follows. Using primers that contained flanking EcoRI and KpnI sites, 3′ portions of the respective genes were amplified by PCR. These sites were then used to insert the amplified fragments into integrative plasmids based on YIplac128 (*LEU2* marked) or YIplac211 (*URA3* marked) (Gietz and Sugino, 1988) that contained sequences encoding epitope tags followed by the terminator sequence of the *CYC1* gene (T_{CYC1}). Each of the resulting plasmids was linearized within the coding sequence for targeted integration into the *S. cerevisiae* genome, yielding strains with one copy of the respective gene (fused in-frame to the tag-coding sequence) expressed from its natural promoter and T_{CYC1}, in addition to a 3′ portion of the same gene without promoter. The epitope tags used were double ha ("ha") and Flag-His$_6$ ("FH") (Dohmen et al., 1995). Plasmid pJD429 (*CEN/URA3*) expressing a *UMP1* cDNA from P*GAL1* was isolated from a cDNA library (Liu et al., 1992) through complementation of the *ump1Δ* mutation.

Pulse-Chase Analyses

R–β-gal (Ub-R–β-gal) and Ub-P–β-gal (Bachmair et al., 1986) were expressed in the *MATa* strains JD47-13C or JD59. MATα2$_{1\text{-}68}$–β-gal was expressed from the plasmid YCp50-α2$_{deg1}$–β-gal (Hochstrasser and Varshavsky, 1990) in the *MATα* strains JD53 and JD81-1A (Table 1). Pulse labeling for 5 min with Redivue Promix [^{35}S] (Amersham) followed by a chase and immunoprecipitation with monoclonal anti–β-gal (Promega) or anti-ha (16B12, Babco) antibodies were carried out as described by Dohmen et al. (1991). ^{35}S proteins fractionated by SDS-PAGE were detected by fluorography.

Fractionation of Whole-Cell Extracts by Gel Filtration

S. cerevisiae were grown at 30°C in YPD or in SD media to OD$_{600}$ of 1.4 ± 0.1, harvested at 3000g, washed with cold water, frozen in liquid nitrogen, and stored at −80°C. Cell paste was ground to powder in a mortar in the presence of liquid nitrogen. The extraction buffer was 50 mM Tris-HCl (pH 7.5), 2 mM ATP, 5 mM MgCl$_2$, 1 mM DTT, 15% glycerol, used at 2 ml per gram of pelleted yeast cells. After centrifugation at 31,000g for 10 min at 2°C, the supernatant was subjected to a second centrifugation at 60,000g for 30 min, yielding an extract with the protein concentration of ~5 mg/ml. The protein concentration of extracts was equalized in parallel experiments, using the extraction buffer. Using the FPLC system (Pharmacia), 200 μl samples of an extract were chromatographed on a Superose-6 column equilibrated with extraction buffer. The flow rate was 0.3 ml/min and fractions of 0.6 ml were collected. The Superose-6 column was calibrated using the following standards: thyroglobulin (669 kDa), ferritin (443 kDa), and bovine serum albumin (66 kDa). Dextran blue was used to monitor the void volume.

Assays for Proteolytic Activities with Fluorogenic Peptide Substrates

To determine the chymotrypsin-like activity, 20 μl of the protein fraction and 20 μl of 0.5 mM succinyl-Leu-Leu-Val-Tyr-7-amido-4-methylcoumarin in 50 mM Tris-HCl (pH 7.8), 2 mM ATP, 5 mM MgCl$_2$, and 1 mM DTT were mixed and incubated for 1 hr at 37°C. The reaction was stopped by addition of 960 μl of cold ethanol, and the fluorescence was measured at 440 nm, using the excitation wavelength of 380 nm. The trypsin-like activity and the peptidylglutamyl peptide-hydrolyzing activity were determined with, respectively, Cbz-Ala-Ala-Arg-4MeO-β-naphthylamide and Cbz-Leu-Leu-Glu-β-naphthylamide as fluorogenic peptide substrates (Fischer et al., 1994), except that the volume of the protein fraction was 75 μl, the reactions were stopped after 270 min at 37°C, and the fluorescence emission was measured at 366 nm, using an excitation wavelength of 420 nm. One unit (U) of a proteolytic activity is defined as 1 μmol of the fluorophore produced per min under these conditions.

Immunoprecipitation, Electrophoresis, and Immunoblotting

Immunoprecipitations were carried out using either monoclonal anti-tag antibodies (see below), or polyclonal rabbit antisera raised against the yeast 20S proteasome (a gift from K. Tanaka), or against Ump1p-His$_6$ expressed in *E. coli*. For immunoblotting, the protein samples were boiled for 3 min in the presence of 2% SDS and 0.1 M 2-mercaptoethanol, then subjected to 12% SDS-PAGE, and thereafter transferred onto a PVDF membrane (Millipore) in a wet blot system (BioRad). The blots were incubated with either rabbit anti-ubiquitin (Ramos et al., 1995), or anti-Cim3p (Ghislain et al., 1993) polyclonal antibodies, or 16B12 anti-ha (BabCo), or M2 anti-Flag (Kodak) monoclonal antibodies, and were processed as described (Ramos et al., 1995), except that the initially blotted proteins were visualized using horseradish peroxidase-conjugated goat anti-mouse or anti-rabbit IgGs, the chemiluminescence blotting substrate detection system from Boehringer Mannheim, and X-ray films. Nondenaturing 4.5% acrylamide gel electrophoresis was performed as described by Hough et al. (1987), and the gels were incubated for 15 min in transfer buffer containing 0.1% SDS before electrotransfer.

Affinity Purification of Proteasomes and Related Complexes

The 20S proteasome and its precursors were purified from strain JD126 that expressed Pre1p-FH and Ump1p-ha. Crude extracts were prepared as described above, followed by an exchange of buffer to 50 mM Na-phosphate (pH 8.0), 0.3 M NaCl, using PD-10 columns (Pharmacia). The extract was then incubated with Ni-NTA agarose (Qiagen) in batch for 2 hr at 4°C, followed by washings and elution according to the manufacturer's protocol, except that the (pH 6.0) washing buffer contained 20 mM imidazole. His$_6$-tagged proteins were eluted with a step gradient of 100-500 mM imidazole. Active fractions containing a mixture of the mature proteasome and its precursors were pooled, diluted 2-fold with 50 mM Tris-HCl (pH 7.5), and incubated for 2 hr at 4°C in batch with 0.5 ml anti-Flag M2 antibody affinity resin (IBI/Eastman Kodak) that had been equilibrated in TN buffer (0.15 M NaCl, 50 mM Tris-HCl, [pH 7.5]). After the loading, the resin was washed with TN buffer. Flag-tagged proteins were specifically eluted with the Flag epitope peptide (IBI/Eastman Kodak).

Acknowledgments

We thank Rohan Baker, Anthony Bretscher, Ricardo Ferreira, Wolfgang Heinemeyer, Wolfgang Hilt, Mark Hochstrasser, Carl Mann, Nancy Kleckner, Keiji Tanaka, and Dieter Wolf for the gifts of plasmids, yeast strains, and antisera; Elisabeth Andrews for assistance in the cloning of *UMP1*, Reiner Stappen for assistance in producing anti-Ump1p antibodies, Robert Kramer (Limberg Druck) for printing the figures, and Nils Johnsson and Ralf Kölling for helpful suggestions. P. C. R., J. H., and R. J. D. are grateful to Cornelis P. Hollenberg for providing lab space and for his support, and to Isabel Fuchs for technical assistance. P. C. R. was supported by a postdoctoral fellowship from Fundação para a Ciência e Tecnologia, Programa Praxis XXI. This work was supported by a grant to R. J. D. from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (0316711), by start-up funding from the Ministerium für

Wissenschaft und Bildung des Landes Nordrhein-Westfalen, and by a grant to A. V. from the National Institutes of Health (GM31530).

Received December 10, 1997; revised January 16, 1998.

References

Alani, E., Cao, L., and Kleckner, N. (1987). A method for gene disruption that allows repeated use of *URA3* selection in the construction of multiply disrupted yeast strains. Genetics *116*, 541–545.

Arendt, C.S., and Hochstrasser, M. (1997). Identification of the yeast 20S proteasome catalytic centers and subunit interactions required for active-site formation. Proc. Natl. Acad. Sci. USA *94*, 7156–7161.

Bachmair, A., Finley, D., and Varshavsky, A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. Science *234*, 179–186.

Baker, R.T., and Varshavsky, A. (1995). Yeast N-terminal amidase—a new enzyme and component of the N-end rule pathway. J. Biol. Chem. *270*, 12065–12074.

Bartel, B., Wünning, I., and Varshavsky, A. (1990). The recognition component of the N-end rule pathway. EMBO J. *9*, 3179–3189.

Chen, P., and Hochstrasser, M. (1996). Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. Cell *86*, 961–972.

Coux, O., Tanaka, K., and Goldberg, A.L. (1996). Structure and functions of the 20S and 26S proteasomes. Annu. Rev. Biochem. *65*, 801–847.

Deveraux, Q., Ustrell, V., Pickart, C., and Rechsteiner, M. (1994). A 26S protease subunit that binds ubiquitin conjugates. J. Biol. Chem. *269*, 7059–7061.

Dohmen, R.J., Madura, K., Bartel, B., and Varshavsky, A. (1991). The N-end rule is mediated by the UBC2(RAD6) ubiquitin-conjugating enzyme. Proc. Natl. Acad. Sci. USA *88*, 7351–7355.

Dohmen, R.J., Wu, P., and Varshavsky, A. (1994). Heat-inducible degron: a method for constructing temperature-sensitive mutants. Science *263*, 1273–1276.

Dohmen, R.J., Stappen, R., McGrath, J.P., Forrova, H., Kolarov, J., Goffeau, A., and Varshavsky, A. (1995). An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. J. Biol. Chem. *270*, 18099–18109.

Fischer, M., Hilt, W., Richter-Ruoff, B., Gonen, H., Ciechanover, A., and Wolf, D.H. (1994). The 26S proteasome of the yeast *Saccharomyces cerevisiae*. FEBS Lett. *355*, 69–75.

Frentzel, S., Pesold-Hurt, B., Seelig, A., and Kloetzel, P.-M. (1994). 20S proteasomes are assembled via distinct precursor complexes. J. Mol. Biol. *236*, 975–981.

Ghislain, M., Udvardy, A., and Mann, C. (1993). *S. cerevisiae* 26S protease mutants arrest cell division in G2/metaphase. Nature *366*, 358–362.

Gietz, R.D., and Sugino, A. (1988). New yeast-*Escherichia coli* shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. Gene *74*, 527–534.

Groll, M., Ditzel, L., Loewe, J., Stock, D., Bochtler, M., Bartunik, H.D., and Huber, R. (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. Nature *386*, 463–471.

Heinemeyer, W., Gruhler, A., Mohrle, V., Mahe, Y., and Wolf, D.H. (1993). *PRE2*, highly homologous to the human major histocompatibility complex-linked *RING10* gene, codes for a yeast proteasome subunit necessary for chymotryptic activity and degradation of ubiquitinated proteins. J. Biol. Chem. *268*, 5115–5120.

Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U., and Wolf, D.H. (1997). The active sites of the eukaryotic 20S proteasome and their involvement in subunit precursor processing. J. Biol. Chem. *272*, 25200–25209.

Hochstrasser, M. (1996). Ubiquitin-dependent protein degradation. Annu. Rev. Genet. *30*, 405–439.

Hochstrasser, M., and Varshavsky, A. (1990). In vivo degradation of a transcriptional regulator: the yeast $\alpha 2$ repressor. Cell *61*, 697–708.

Hochstrasser, M., Papa, F.R., Chen, P., Swaminathan, S., Johnson, P., Stillman, L., Amerik, A.Y., and Li, S.J. (1995). The doa pathway—studies on the functions and mechanisms of ubiquitin-dependent protein degradation in the yeast *Saccharomyces cerevisiae*. Cold Spring Harb. Symp. Quant. Biol. *60*, 503–513.

Hough, R., Ratt, G., and Rechsteiner, M. (1987). Purification of two high molecular weight proteases from rabbit reticulocyte lysate. J. Biol. Chem. *262*, 8303–8313.

Jentsch, S., and Schlenker, S. (1995). Selective protein degradation—a journey's end within the proteasome. Cell *82*, 881–884.

Johnson, E.S., Ma, P.C., Ota, I.M., and Varshavsky, A. (1995). A proteolytic pathway that recognizes ubiquitin as a degradation signal. J. Biol. Chem. *270*, 17442–17456.

Jungmann, J., Reins, H.A., Schobert, C., and Jentsch, S. (1993). Resistance to cadmium mediated by ubiquitin-dependent proteolysis. Nature *361*, 369–371.

Liu, H., Krizek, J., and Bretscher, A. (1992). Construction of a GAL1-regulated yeast cDNA expression library and its application to the identification of genes whose overexpression causes lethality in yeast. Genetics *132*, 665–673.

Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W., and Huber, R. (1995). Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. Science *268*, 533–539.

Lupas, A., Flanagan, J.M., Tamura, T., and Baumeister, W. (1997). Self-compartmentalizing proteases. Trends Biochem. Sci. *22*, 399–404.

Nandi, D., Woodward, E., Ginsburg, D.B., and Monaco, J. (1997). Intermediates in the formation of mouse 20S proteasomes: implications for the assembly of precursor β subunits. EMBO J. *16*, 5363–5375.

Peters, J.M. (1994). Proteasomes: protein degradation machines of the cell. Trends Biochem. Sci. *19*, 377–382.

Ramos, P., Cordeiro, A., Ferreira, R., Ricardo, C., and Teixeira, A. (1995). The presence of ubiquitin-protein conjugates in plant chloroplast lysates is due to cytosolic contamination. Austr. J. Plant Physiol. *22*, 893–901.

Rose, M.D., Novick, P., Thomas, J.H., Botstein, D., and Fink, G.R. (1987). A *Saccharomyces cerevisiae* genomic plasmid bank based on a centromere-containing shuttle vector. Gene *60*, 237–243.

Rothstein, R. (1991). Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol. *194*, 281–301.

Schmidtke, G., Schmidt, M., and Kloetzel, P.-M. (1997). Maturation of mammalian 20S proteasome: purification and characterization of 13S and 16S proteasome precursor complexes. J. Mol. Biol. *268*, 95–106.

Seemüller, E., Lupas, A., Stock, D., Löwe, J., Huber, R., and Baumeister, W. (1995). Proteasome from *Thermoplasma acidophilum*: a threonine protease. Science *268*, 579–582.

Seemüller, E., Lupas, A., and Baumeister, W. (1996). Autocatalytic processing of the 20S proteasome. Nature *382*, 468–470.

Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics *122*, 19–27.

Varshavsky, A. (1996). The N-end rule—functions, mysteries, uses. Proc. Nat. Acad. Sci. USA *93*, 12142–12149.

Varshavsky, A. (1997). The ubiquitin system. Trends Biochem. Sci. *22*, 383–387.

Yang, Y., Früh, K., Ahn, K., and Peterson, P.A. (1995). In vivo assembly of the proteasomal complexes, implications for antigen processing. J. Biol. Chem. *270*, 27687–27694.

ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS
Vol. 357, No. 2, September 15, pp. 317–325, 1998
Article No. BB980829

The N-End Rule Pathway in *Xenopus* Egg Extracts

Ilia V. Davydov, Debabrata Patra, and Alexander Varshavsky[1]
Division of Biology, California Institute of Technology, Pasadena, California 91125

Received April 7, 1998, and in revised form June 30, 1998

Ubiquitin-dependent degradation of intracellular proteins underlies a multitude of biological processes, including the cell cycle, cell differentiation, and responses to stress. One ubiquitin-dependent proteolytic system is the N-end rule pathway, whose targets include proteins that bear destabilizing N-terminal residues. This pathway, which has been characterized only in somatic cells, is shown here to be present also in germ line cells such as the eggs of the amphibian *Xenopus laevis.* We demonstrate that the set of destabilizing residues in the N-end rule pathway of *Xenopus* eggs is similar, if not identical, to that of somatic cells such as mammalian reticulocytes and fibroblasts. It is also shown that the degradation of engineered N-end rule substrates in egg extracts can be strongly and selectively inhibited by dipeptides bearing destabilizing N-terminal residues. This result allowed us to ask whether selective inhibition of the N-end rule pathway in egg extracts influences the apoptosis-like changes that are observed in these extracts. A dipeptide bearing a bulky hydrophobic (type 2) destabilizing N-terminal residue was found to delay the apoptotic changes in egg extracts, whereas dipeptides bearing basic (type 1) destabilizing N-terminal residues had no effect. High activity of the N-end rule pathway in egg extracts provides an alternative to reticulocyte extracts for the *in vitro* analyses of this pathway.

© 1998 Academic Press

Key Words:* ubiquitin; proteolysis; N-end rule; apoptosis; germ line; *Xenopus.

A number of regulatory circuits, including those that control the cell cycle, cell differentiation, and responses to stress, involve metabolically unstable proteins (1–4). A short *in vivo* half-life of a regulator provides a way to generate its spatial gradients and allows for rapid adjustments of its concentration, or subunit composition, through changes in the rate of its synthesis or degradation. Damaged or otherwise abnormal proteins tend to be short-lived as well.

Features of proteins that confer metabolic instability are called degradation signals, or degrons (5). The essential component of one degradation signal, called the N-degron, is a destabilizing N-terminal residue of a protein (6, 7). The set of amino acid residues which are destabilizing in a given cell yields a rule, called the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar but distinct versions of the N-end rule pathway are present in all organisms examined, from mammals to fungi and bacteria (7).

In eukaryotes, an N-degron comprises at least two determinants: a destabilizing N-terminal residue and an internal lysine or lysines (8–10). The Lys residue is the site of formation of a multiubiquitin chain (11). The N-end rule pathway is thus a part of the ubiquitin (Ub)[2] system. Ub is a 76-residue eukaryotic protein whose covalent conjugation to other proteins plays a role in a multitude of biological processes (1, 12–15). In most of them, Ub acts through routes that involve the degradation of Ub-protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease (16–18).

It was found that linear Ub fusions are rapidly cleaved by Ub-specific proteases (UBPs) at the Ub-protein junction, making possible the production of otherwise identical proteins bearing different N-terminal residues. This technical advance led to the discovery of the N-end rule pathway (6, 7). The N-end rule is organized hierarchically. In the yeast *Saccharomyces cerevisiae,* Asn and Gln are *tertiary* destabilizing N-terminal residues in that they function through their conversion, by enzymatic deamidation, into the *second-*

---

[1] To whom correspondence should be addressed at Division of Biology, 147-75, Caltech, 1200 East California Boulevard, Pasadena, CA 91125. Fax: (626) 440-9821. E-mail: avarsh@cco.caltech.edu.

[2] Abbreviations used: Ub, ubiquitin; DHFR, mouse dihydrofolate reductase; R-transferase, Arg-tRNA-protein transferase; CSF, cytostatic factor; TCA, trichloroacetic acid; DTT, dithiothreitol; AMC, 7-amino-methylcoumarin; UBP, Ub-specific processing protease.

0003-9861/98 $25.00
Copyright © 1998 by Academic Press
All rights of reproduction in any form reserved.

ary destabilizing N-terminal residues Asp and Glu. The destabilizing activity of N-terminal Asp and Glu requires their conjugation, by Arg-tRNA-protein transferase (R-transferase), to Arg, one of the *primary* destabilizing residues (19–21). The primary destabilizing N-terminal residues are bound directly by N-recognin (also called E3), the recognition component of the N-end rule pathway. In *S. cerevisiae,* N-recognin is a 225-kDa protein, encoded by *UBR1,* that targets potential N-end rule substrates through the binding to their primary destabilizing N-terminal residues — Phe, Leu, Trp, Tyr, Ile, Arg, Lys, or His (7, 22). N-recognin has at least two substrate-binding sites. The type 1 site is specific for the basic N-terminal residues Arg, Lys, and His. The type 2 site is specific for the bulky hydrophobic N-terminal residues Phe, Leu, Trp, Tyr, and Ile (7).

The currently known physiological substrates of the N-end rule pathway are the RNA polymerase of Sindbis virus (23), the *GPA1*-encoded Gα subunit of the *S. cerevisiae* heterotrimeric G protein (24), and Cup9p, a short-lived transcriptional repressor in *S. cerevisiae* that controls the expression of *PTR2,* which encodes a peptide transporter (25). The Ubr1p-Cup9p-Ptr2p circuit, which controls the import of peptides in yeast, is the first clear instance of a physiological function of the N-end rule pathway (25). Among the cells of multicellular organisms, this proteolytic system was characterized in rabbit reticulocytes (19) and L cells, a line of fibroblast-like mouse cells (26). Recently, it was suggested that the N-end rule pathway might play a role in apoptosis (programmed cell death) (7). A way to verify this conjecture would be to use a mutant that lacks the N-end rule pathway. Such mutants were constructed and characterized in *Escherichia coli* and *S. cerevisiae* (7) but not yet in multicellular organisms. Therefore we considered the use of N-end rule inhibitors.

Previous work has shown that the addition of amino acid derivatives such as dipeptides that bear destabilizing N-terminal residues to reticulocyte extract results in a strong and selective inhibition of the N-end rule pathway in the extract (19, 27). Specifically, dipeptides bearing type 1 destabilizing N-terminal residues inhibited the degradation of test N-end rule substrates bearing basic (type 1) destabilizing N-terminal residues but had no effect on the degradation of substrates bearing type 2 N-terminal residues. A converse pattern was observed with dipeptides bearing bulky hydrophobic (type 2) destabilizing N-terminal residues (19). Dipeptides added to *S. cerevisiae* cultures have been demonstrated to inhibit the N-end rule pathway *in vivo* as well (28). However, experiments with the N-end rule pathway of intact mammalian cells have shown dipeptides to be ineffective inhibitors in this setting[3].

Given this latter constraint, we sought to explore the physiological consequences of inhibiting a metazoan N-end rule pathway in the inhibitor-accessible setting of a cell-free system. We also wished to determine whether the N-end rule pathway is present in germ line cells such as the eggs of the amphibian *Xenopus laevis,* a major experimental organism in studies of embryogenesis and cell cycle control (4, 29). Our findings are described below.

MATERIALS AND METHODS

Construction of plasmids. The plasmids pT7-Ub**M**e^{K}DHFRhis and pT7-Ub**R**e^{K}DHFRhis expressed Ub-Met-e^{K}-DHFR-His$_6$ and Ub-Arg-e^{K}-DHFR-His$_6$ fusions (denoted, respectively, as Ub-Met-DHFR and Ub-Arg-DHFR) in *E. coli* from the T7 polymerase promoter (see Results for the definitions of e^{K} and other terms). These plasmids were constructed from pEJJ1-M and pEJJ1-R (30). The *Kpn*I-*Hin*dIII fragment of pEJJ1-M and pEJJ1-R was replaced with a synthetic oligonucleotide duplex (5′-CCATCACCATCACCATCACTAAA-3′ and 5′-AGCTTTTAGTGATGGTGATGGTGATGGGTAC-3′) that encoded the His$_6$ tag and bore the *Kpn*I and *Hin*dIII overhangs. The other pT7-Ub**X**e^{K}DHFRhis plasmids, which encoded otherwise identical Ub-X-DHFR fusions bearing different X residues (Met, Arg, Leu, Phe, Cys, Asp, or Asn) (Fig. 1), were constructed from pT7-Ub**R**e^{K}DHFRhis by site-directed mutagenesis, using PCR (31).

Overexpression, labeling, and purification of Ub-X-e^{K}-DHFR-His$_6$ proteins. A pT7-Ub**X**e^{K}DHFRhis plasmid was introduced into *E. coli* BL21 (DE3) (31). Protein expression was induced in *E. coli* by adding isopropyl-1-thio-β-*D*-galactopyranoside (IPTG), and cells were labeled with [^{35}S]methionine/cysteine (^{35}S-Express, New England Nuclear, Boston, MA), as described (30). The labeled cells were collected by centrifugation and disrupted by sonication, and a ^{35}S-labeled Ub-X-e^{K}-DHFR-His$_6$ (Ub-X-DHFR) test protein was purified by affinity chromatography under nondenaturing conditions, using the Ni-NTA Spin Kit (Qiagen, Chatsworth, CA), and dialyzed against 1 mM MgCl$_2$, 1 mM dithiothreitol (DTT), 0.1 M Tris–HCl (pH 7.7), frozen rapidly, and stored at −80°C in samples that were to be thawed only once. The specific radioactivity of [^{35}S]Ub-X-DHFR proteins was 5–10 $\times$ 10^5 cpm/μg.

Degradation assays in Xenopus egg extracts. Cytostatic factor (CSF)-arrested egg extracts were prepared from *Xenopus* eggs as described (32). In most experiments, cycloheximide (0.1 mg/ml) was added to the extracts to preclude reincorporation of [^{35}S]methionine into newly made proteins. In some experiments (see the legends to figures), egg extracts were activated by the addition of 0.4 mM CaCl$_2$ 1 h before the assay. [^{35}S]Ub-X-DHFR test proteins were added to the extract to the final concentration of ~25 μg/ml. Dipeptides were added together with bestatin (Sigma, St. Louis, MO) to the final concentrations of 10 mM and 50 μg/ml, respectively. Bestatin was added to decrease the degradation of dipeptides in the extract (27). (Control experiments (not shown) showed that the addition of bestatin alone did not significantly inhibit the degradation of test proteins by the N-end rule pathway.) The following dipeptides and other amino acid derivatives were used: Arg-β-Ala, Ala-Lys, Lys-Ala, Trp-Ala (Sigma), and Tyr-His (Bachem Science, King of Prussia, PA). Stock samples of dipeptides were 0.5 M solutions in 10 mM K-Hepes, pH 7.5.

To follow the degradation of test proteins, reaction mixtures (0.1 ml) prepared as described above were incubated at 23°C. Samples (2.5 μl) were withdrawn in duplicate for each time point. One sample was examined by SDS-12% PAGE and autoradiography, using PhosphorImager (Molecular Dynamics, Sunnyvale, CA) (33). The other sample was used to determine the relative amount (%) of ^{35}S soluble in 5% trichloroacetic acid (TCA) (30). This parameter was calculated as follows:

---

[3] F. Lévy and A. Varshavsky, unpublished data.

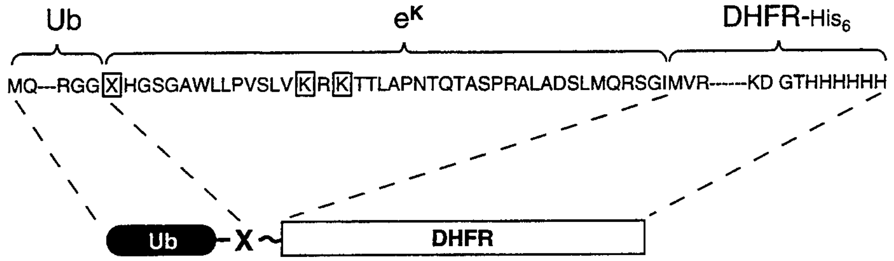

Ub
eK
DHFR-His6
MQ---RGG[X]HGSGAWLLPVSLV[K]R[K]TTLAPNTQTASPRALADSLMQRSGIMVR------KD GTHHHHHH
Ub
X
DHFR

FIG. 1. The Ub-X-DHFR test proteins. These fusions, identical except for the variable residue X adjacent to the Ub moiety, were expressed and radiolabeled in *E. coli*, purified, and used as described under Materials and Methods. X was either Met, Arg, Leu, Phe, Cys, Asp, or Asn. The *E. coli* Lac repressor-derived 42-residue sequence, denoted as e^K (extension (e) containing lysines (K)) was described previously (8, 30, 37). The alternative ubiquitylation sites Lys-15 and Lys-17 are boxed in the sequence of e^K (8). DHFR represents DHFR-His$_6$, the mouse DHFR whose C-terminus was extended by 8 residues that included the His$_6$ tag.

$$\text{percent TCA-soluble } ^{35}\text{S} = \frac{X}{Y} \cdot \frac{a}{a-1} \cdot 100\%,$$

where X is the amount of TCA-soluble ^{35}S (cpm); Y is the total amount of ^{35}S (cpm) in the same sample; a is the number of Met residues in a Ub-X-DHFR test protein (a = 10 for Ub-Met-DHFR; a = 9 for the other Ub-X-DHFRs). The $a - 1$ term above corrects for the presence of one Met residue in Ub.

The data presented in each of the experimental Figures were produced using the same extract on the same day. Unless stated otherwise, the data were consistent between at least two (usually three or more) independent experiments carried out with different preparations of the extract.

In vitro apoptosis assays. Apoptotic extracts from *Xenopus* eggs were prepared according to Newmeyer *et al.* (34). The difference between control and apoptotic egg extracts was in the time intervals between hormone injections used to produce the eggs. Frogs were injected with 75 units of pregnant mare serum gonadotropin (Calbiochem, San Diego, CA) 14–30 days (in contrast to 3–10 days for control extracts) before they were induced to lay eggs by injecting 800 units of human chorionic gonadotropin (Sigma). Eggs were collected, and the extract was prepared as described (32). The extracts were used within 1 h after preparation. For most assays, the extracts were arrested in interphase by the addition of 0.4 mM CaCl$_2$ and 0.1 mg/ml cycloheximide. Demembranated *Xenopus* sperm nuclei (32), at ~1000 per μl of extract, were also added at this time. Extracts treated as described were distributed into 0.1-ml samples and incubated at 23°C. Apoptotic changes in the extracts were monitored in two ways: by observing shrinkage and disintegration of the added sperm nuclei using phase contrast and fluorescent microscopy (34) and by measuring the DEVD-specific protease activity. The latter assay determined the total activity of caspases that recognize the tetrapeptide sequence DEVD (Asp-Glu-Val-Asp) (35, 36).

The DEVD-specific protease activity was measured by incubating 5 μl of extract with a quenched fluorescent substrate Ac-DEVD-AMC (50 μM) (Bachem Bioscience) in 0.1 ml of the extract buffer (15 mM MgCl$_2$, 1 mM DTT, 20 mM EGTA, 80 mM Na-β-glycerolphosphate, pH 7.3). The reaction was carried out for 10 min at 23°C, followed by freezing of a 20-μl sample in liquid N$_2$. Samples were thawed by diluting into 1 ml of phosphate-buffered saline, and their fluorescence (excitation at 380 nm and emission at 460 nm) was measured in the Hitachi F-4500 spectrofluorimeter (35, 36). AMC was used as a fluorescence standard.

RESULTS

The N-End Rule Pathway in Xenopus Egg Extracts

Among the cells of multicellular organisms, the N-end rule pathway has been identified and analyzed in rabbit reticulocytes (19) and L cells, a line of transformed, fibroblast-like mouse cells (26). To determine whether the N-end rule pathway is present in germ line cells such as *Xenopus* eggs, we used an approach similar to the earlier one with rabbit reticulocytes (19, 30).

A set of N-end rule substrates differing exclusively by their N-terminal residues was constructed, expressed in *E. coli*, metabolically labeled with [^{35}S]methionine, and purified by affinity chromatography (see Materials and Methods). A test protein of this set, Ub-X-e^K-DHFR-His$_6$, denoted below as Ub-X-DHFR, contained the following parts: the N-terminal Ub moiety; a variable residue X; e^K, a 42-residue, *E. coli* Lac repressor-derived sequence that contained the second (lysine) determinant of the N-degron (8, 37); and the His$_6$-tagged mouse dihydrofolate reductase moiety (Fig. 1). Ub fusions are not cleaved at the Ub-protein junction in *E. coli*, which lacks the Ub system (38). By contrast, in eukaryotes the Ub fusions, including the Ub-X-DHFRs, are rapidly cleaved by UBPs after the last residue of Ub, making it possible to produce, *in vivo* or *in vitro*, otherwise identical test proteins such as X-DHFRs that differ exclusively by their N-terminal residues (6, 7).

CSF-arrested *Xenopus* egg extracts were prepared from unfertilized frog eggs (32). Specific ^{35}S-labeled Ub-X-DHFR test proteins (Fig. 1) were added to the egg extracts, and their metabolic fates were monitored either by SDS–PAGE and autoradiography or by measuring the amount of TCA-soluble ^{35}S released during incubation at 23°C. The identity of residue X in the tested X-DHFRs encompassed at least one representative of each class of N-terminal destabilizing residues: Asn (tertiary destabilizing); Asp and Cys (secondary destabilizing), Arg (type 1 primary destabilizing), Phe and Leu (type 2 primary destabilizing), and Met (stabilizing) (see introduction) (7).

As expected, a Ub-X-DHFR test protein was deubiquitylated upon its addition to the extract. This was detected through an increase in the electrophoretic mobility of the major ^{35}S-labeled species and the con-

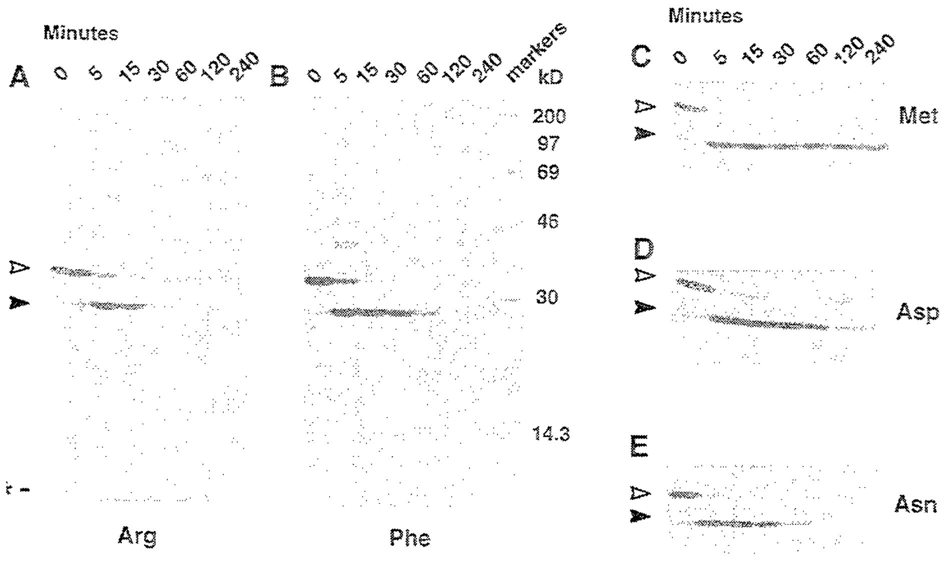


FIG. 2. Deubiquitylation and degradation of Ub-X-DHFR test proteins in *Xenopus* egg extract. (A) ^{35}S-labeled, purified Ub-Arg-DHFR was added to a CSF-arrested *Xenopus* egg extract and incubated at 23°C for the indicated times. The samples were analyzed by SDS–PAGE, and the labeled species were detected by autoradiography (see Materials and Methods). (B) Same as A, but with Ub-Phe-DHFR. (C) Same as A, but with Ub-Met-DHFR (only the area around Met-DHFR is shown). (D) Same as C, but with Ub-Asp-DHFR. (E) Same as C, but with Ub-Asn-DHFR. Open arrowheads indicate the bands of the initial, 34-kDa Ub-X-DHFR. Closed arrowheads indicate the bands of deubiquitylated, 26-kDa X-DHFR. An asterisk indicates the band of free [^{35}S]Ub, which migrated close to the dye front in this electrophoretic system. The molecular masses (in kDa) of protein markers are indicated to the right of B.

comitant appearance of a labeled ~8 kDa species (Fig. 2 and data not shown). Deubiquitylation of Ub-X-DHFR in the extract was essentially complete in 5 min or less (Fig. 2). Whereas the relative intensity of the free Ub band remained nearly constant during the subsequent 4-h incubation (Figs. 2A and 2B, and data not shown), the relative intensity of the X-DHFR band was either constant or decreased, at a varying rate, depending on the identity of the N-terminal residue X. Based on their relative metabolic stabilities in the extract (Figs. 2 and 3A, and data not shown), the X-DHFR test proteins could be reproducibly ranked from the long-lived to the short-lived as follows: Met > Cys > Leu > Asp > Asn > Phe > Arg. We conclude that *Xenopus* egg extracts contain an N-end rule pathway whose rule book is similar, and may be identical, to those found in rabbit reticulocytes (19) and mouse L cells (26).

The decay curves of X-DHFRs that were determined by monitoring the release of TCA-soluble ^{35}S and the resulting ranking of their N-terminal destabilizing residues were reproducible between experiments that utilized independent preparations of the egg extract (data not shown). Specifically, it took ~10 min to destroy 50% of Arg-DHFR to acid-soluble fragments. By 30 min, the degradation of Arg-DHFR was ~90% complete (Figs. 2A and 3A).

Approximately 20% of the added Met-DHFR, which bore a stabilizing N-terminal residue (7), was degraded during the first 10 min of incubation, and another 5 to 15% was degraded over the next 30 min. However, the rest of the Met-DHFR remained stable in the extract (Figs. 2C and 3A). A likely explanation of this result is that our preparation contained two types of Met-DHFR molecules: ~70% were undamaged and long-lived in the extract, as observed with Met-DHFR *in vivo,* in both *S. cerevisiae* and mouse L cells (Ref. 8; F. Lévy and A. V., unpubl. data), whereas the remainder comprised misfolded or otherwise damaged Met-DHFR molecules that were degraded by a proteolytic system distinct from the N-end rule pathway. By inference, the same background of degradation of a subpopulation of molecules would be expected for other X-DHFRs. It is unclear whether a subpopulation of misfolded or otherwise damaged X-DHFRs resulted from their overexpression in *E. coli,* or whether the damage occurred largely during purification of X-DHFRs. Independent evidence for this explanation of the background degradation of Met-DHFR is presented below.

Previous work has shown that N-terminal Cys is a stabilizing residue in *S. cerevisiae* (8), but is a secondary destabilizing residue in rabbit reticulocyte extracts and mouse L cells (19, 26). In *Xenopus* egg extracts, Cys was found to be a weakly destabilizing residue. Specifically, among the tested X-DHFRs, Cys-DHFR was the longest-lived test protein save for Met-DHFR, which bore a stabilizing N-terminal residue (Fig. 3A). Substrates that bear secondary or tertiary destabilizing residues require the tRNA-dependent activity of R-transferase for their degradation by the N-end rule

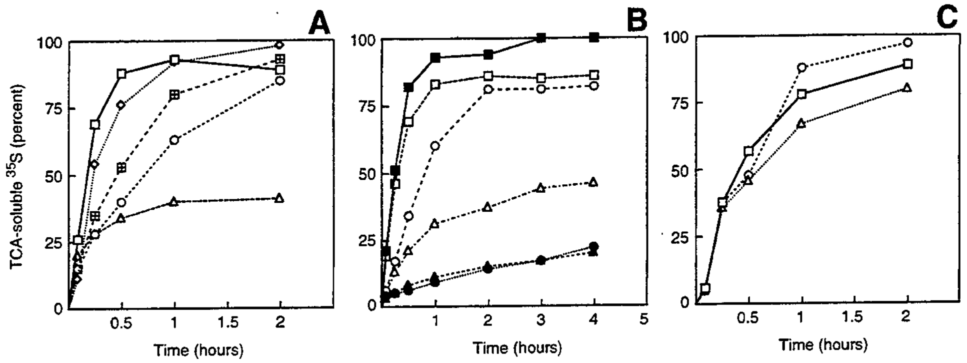


FIG. 3. Degradation of X-DHFR test proteins in egg extracts under different conditions, as determined by measuring TCA-soluble ^{35}S. (A) ^{35}S-labeled, purified Ub-Arg-DHFR (□), Ub-Phe-DHFR (◇), Ub-Met-DHFR (△), Ub-Asp-DHFR (○), and Ub-Asn-DHFR (⊞) were added to samples of the CSF-arrested *Xenopus* egg extract. After incubation at 23°C for the indicated times, the amounts of ^{35}S soluble in 5% TCA were determined for each time point (see Materials and Methods). Since the free [^{35}S]Ub, produced through deubiquitylation of a Ub-X-DHFR, was stable during the incubation, 100% of TCA-soluble ^{35}S represented complete degradation of the Ub-lacking [^{35}S]X-DHFR (see Materials and Methods for further details). (B) Pretreatment of egg extract with RNAase selectively inhibits degradation of N-end rule substrates bearing secondary destabilizing N-terminal residues. A CSF-arrested extract was pretreated with 5 units/ml of RNAase A-agarose (Sigma) for 1 h at 23°C, followed by the removal of RNAase-agarose by centrifugation (*closed symbol* curves). Another sample of the same extract was incubated for 1 h without RNAase (*open symbol* curves). ^{35}S-labeled Ub-Cys-DHFR (*triangles*), Ub-Asp-DHFR (*circles*), or Ub-Arg-DHFR (*squares*) were added to both extracts, and the relative amounts of TCA-soluble ^{35}S released during the incubation were determined as in A. (C) Different states of egg extract do not significantly alter the activity of the N-end rule pathway. ^{35}S-labeled Ub-Phe-DHFR was added to the CSF-arrested egg extract (□), to the CaCl$_2$-treated extract that proceeded toward mitosis (△), and to the CaCl$_2$- and cycloheximide-treated, interphase-arrested extract (○). The degradation of Phe-DHFR was monitored as described in A.

pathway (see introduction). Previous work has shown that pretreatment of reticulocyte extract with RNAase A does not significantly affect the degradation of N-end rule substrates bearing primary destabilizing residues but abolishes the degradation of otherwise identical substrates bearing secondary or tertiary destabilizing N-terminal residues (19). In agreement with these findings, preincubation of *Xenopus* egg extract with RNAase A selectively inhibited the degradation of Cys-DHFR and Asp-DHFR, but had no effect on the degradation of Arg-DHFR (Fig. 3B). We conclude that both Cys and Asp are secondary destabilizing residues in the N-end rule pathway of *Xenopus* eggs. Taken together with the earlier findings about the N-end rule pathway in reticulocyte extract (19), the present findings are consistent with the conjecture that the set of secondary and tertiary destabilizing residues in the N-end rule of *Xenopus* eggs is identical to that of rabbit reticulocytes: two tertiary destabilizing residues (Asn and Gln) and three secondary ones (Asp, Glu, and Cys).

We also asked whether the rate of degradation of an N-end rule substrate depends on physiologically relevant changes in the state of an egg extract. Unfertilized *Xenopus* eggs are arrested at the metaphase of meiosis II through the action of cytostatic factor (39, 40). Fertilization results in a transient increase of the Ca^{2+} concentration in the eggs that inactivates CSF, allowing cells to enter interphase and proceed to mitosis. This process can be mimicked *in vitro* by the addition of Ca^{2+} (32). Such extracts, referred to as cycling extracts, undergo changes characteristic of eggs at mitosis, as observed through alterations in the morphology of *Xenopus* sperm nuclei added to the extract. The addition of cycloheximide and CaCl$_2$ to a CSF-arrested extract results not only in the escape from CSF arrest but also in a subsequent arrest at interphase, owing to the inhibition of protein synthesis, specifically the synthesis of cyclins (32). We examined the degradation of Phe-DHFR, bearing a type 2 primary destabilizing N-terminal residue, in the CSF-arrested, interphase-arrested, and cycling egg extracts, and found no significant differences in the rate of degradation of this N-end rule substrate in different extracts (Fig. 3C).

Dipeptides Bearing Destabilizing N-Terminal Residues Are Efficacious Inhibitors of the N-End Rule Pathway in Xenopus Egg Extracts

Previous work has shown that the addition of amino acid derivatives such as dipeptides bearing destabilizing N-terminal residues to reticulocyte extracts or intact yeast cells results in a strong and selective inhibition of the N-end rule pathway in these settings (19, 27). However, experiments with intact mammalian cells in culture have shown dipeptides to be at most weak inhibitors of the N-end rule pathway in this setting[3]. The present work stemmed in part from the possibility of exploring the physiological consequences of perturbing a metazoan N-end rule pathway in the inhibitor-accessible setting of a cell-free system such as a *Xenopus* egg extract.

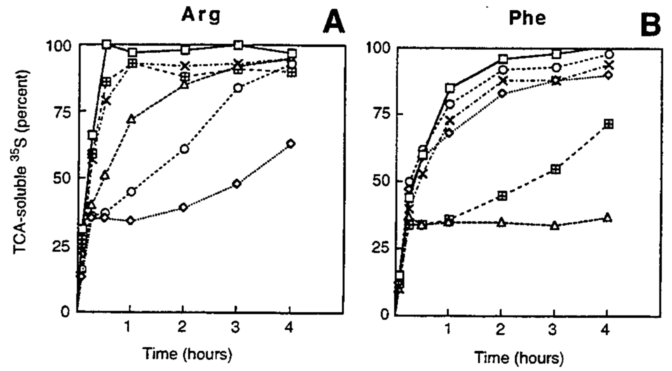


FIG. 4. Inhibition of the N-end rule pathway in egg extracts by dipeptides bearing destabilizing N-terminal residues. (A) ^{35}S-labeled, purified Ub-Arg-DHFR was added to CSF-arrested *Xenopus* egg extracts in the absence (□) or the presence of the following dipeptides (10 mM), together with bestatin (50 μg/ml): Arg-β-Ala (◇), Lys-Ala (○), Tyr-His (△), Trp-Ala (⊞), and Ala-Lys (X). After incubation at 23°C for the indicated times, the amounts of ^{35}S soluble in 5% TCA were determined for each time point. (B) Same as in A but with Ub-Phe-DHFR.

As shown in Figs. 4 and 5, Lys-Ala and especially Arg-β-Ala, both of which bear a type 1 primary destabilizing N-terminal residue, strongly inhibited the degradation of Arg-DHFR in the egg extract. This inhibition was selective, in that the same dipeptides had no significant effect on the degradation of Phe-DHFR, which bore a type 2 primary destabilizing N-terminal residue. Conversely, Tyr-His and Trp-Ala, which bear type 2 destabilizing N-terminal residues, strongly inhibited the degradation of Phe-DHFR, but had at most a small effect on the degradation of Arg-DHFR (Figs. 4 and 5). Ala-Lys, which bears a type 3 destabilizing N-terminal residue (19), did not inhibit the degradation of either Arg- or Phe-DHFRs (Fig. 4).

The dipeptides had little effect on the degradation of their cognate N-end rule substrates during the first 15 min of the incubation, but subsequent proteolysis was strongly inhibited (Fig. 4). This result is consistent with the finding that 20–30% of Met-DHFR, which bears a stabilizing N-terminal residue, was rapidly degraded, whereas the rest of the added Met-DHFR was stable (Fig. 3A). As described above, the latter finding could be accounted for if our preparation of Met-DHFR (and, by inference, of other X-DHFRs) comprised two distinct populations of molecules: undamaged ones (~70%) and hence, in the case of Met-DHFR, long-lived in the extract, and misfolded or otherwise damaged molecules that were recognized and degraded by a proteolytic system distinct from the N-end rule pathway. The failure of dipeptides to inhibit the initial burst of degradation of their cognate X-DHFRs (Fig. 4) is in agreement with this explanation.

Apoptosis and the N-End Rule Pathway

Several groups have described the use of *Xenopus* egg extracts to analyze the process of apoptosis (34, 36, 41–43). After several hours of incubation of the interphase-arrested, apoptotic egg extracts, the activity of DEVD-specific caspases (a subgroup in the ICE/CED-3 family of cysteine proteases) (43) rises sharply, followed by fragmentation of the added sperm nuclei (34). Both the induction of caspases and the fragmentation of the added nuclei in the egg extract can be suppressed by the addition of Bcl2, a protein that inhibits apoptosis *in vivo* (34, 36). Recent work has shown that apoptotic changes in egg extract are triggered, through unknown cytosolic factors, by the release of cytochrome *c* from the mitochondrial fraction of the extract (43). Apparently Bcl2 inhibits apoptosis by blocking the release of cytochrome *c* from mitochondria (43, 44).

We wished to determine whether the apoptosis-like events in egg extracts would be perturbed by dipeptide inhibitors of the N-end rule pathway. In preliminary experiments, we found that the times of apoptotic changes in egg extracts varied considerably between independent extract preparations. Specifically, the sharp rise of the DEVD-specific protease activity and the disintegration of the added sperm nuclei could occur as early as 2 h after the start of incubation or as

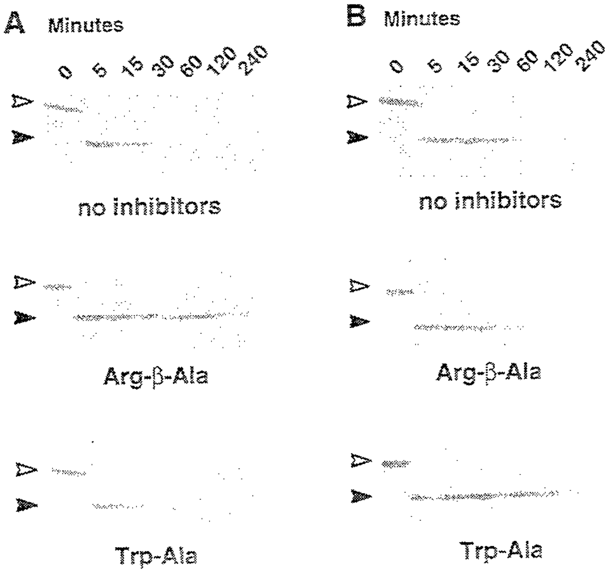


FIG. 5. Electrophoretic analysis of N-end rule substrates in egg extracts in the presence of dipeptides. (A) ^{35}S-labeled, purified Ub-Arg-DHFR was added to a CSF-arrested *Xenopus* egg extract in the absence or the presence of dipeptides Arg-β-Ala or Trp-Ala (at 10 mM, together with bestatin at 50 μg/ml). After incubation at 23°C for the indicated times, the samples were analyzed by SDS–PAGE, and the labeled proteins were detected by autoradiography. Open arrowheads indicate the bands of the initial, 34-kDa Ub-Arg-DHFR. Closed arrowheads indicate the bands of deubiquitylated, 26-kDa Arg-DHFR. (B) Same as in A but with Ub-Phe-DHFR.

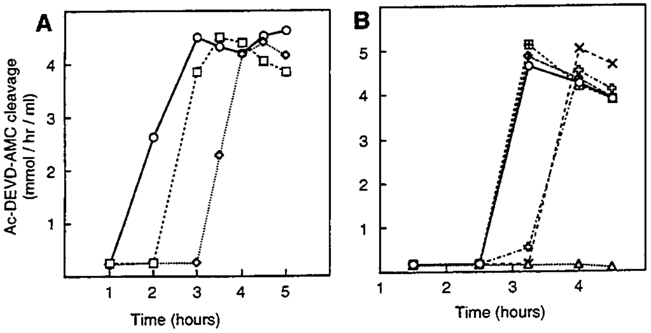


FIG. 6. Apoptotic changes in *Xenopus* egg extracts, assayed by measuring the DEVD-specific protease activity. (A) A preparation of the CSF-arrested egg extract was divided into three samples. The extracts were left untreated (□) or were activated by either 0.4 mM $CaCl_2$ (◇) or 0.4 mM $CaCl_2$ in the presence of 0.1 mg/ml cycloheximide (○). The samples were incubated at room temperature for up to 5 h. The DEVD-specific protease activity in the extracts was determined using the fluorogenic substrate Ac-DEVD-AMC. (B) Samples of interphase-arrested egg extract were incubated either in the absence (○) or in the presence of the following reagents: Lys-Ala at 20 mM (⊞); Ala-Lys at 20 mM (◆); Trp-Ala at 20 mM (×); Trp-Ala plus Lys-Ala, each at 10 mM (✣); Ac-DEVD-CHO (an inhibitor of DEVD proteases, at 10 mM) (△). The dipeptides were added together with bestatin (50 μg/ml). The DEVD-specific protease activity was measured as in A.

late as 5 h (data not shown). This variability, reported by others as well (43), is likely to result in part from the variability of the timing of upstream events that trigger the beginning of apoptosis in the *Xenopus* egg extract. In addition, the timing of apoptotic changes depended on the state of the extracts: when a single preparation was split into three samples, and was used as CSF-arrested, cycling, and interphase-arrested extracts, the increase of the DEVD-specific protease activity was observed first in the interphase-arrested extract, then in the CSF-arrested one, and later in the cycling extract (Fig. 6A). This pattern was reproducible between different preparations of the extract, even though the absolute timing of the rise of the DEVD-specific protease activity varied significantly between preparations (data not shown).

We asked whether selective inhibition of the N-end rule pathway would affect the nature or kinetics of the apoptotic changes. Dipeptides that had been shown to be efficacious inhibitors of the N-end rule pathway (Figs. 2 and 3A) were added to the interphase-arrested and cycling extracts at the beginning of incubation. We found that Lys-Ala and Arg-β-Ala, containing type 1 primary destabilizing N-terminal residues, and Ala-Lys, containing a type 3 destabilizing residue, did not alter the timing of DEVD-specific protease induction (Fig. 6B). In contrast, Trp-Ala, which bears a type 2 destabilizing N-terminal residue, delayed both the rise of the DEVD-specific protease activity and the disintegration of sperm nuclei by ~1 h in 8 out of 12 independent experiments (Fig. 6B and data not shown). The ~1 h delay of apoptotic changes by Trp-Ala was significant but weak in comparison to the effect of Ac-DEVD-CHO, an inhibitor of DEVD-specific proteases, which delayed apoptosis by at least 2 h (data not shown). Interestingly, the proteasome inhibitor 5-iodo-4-hydroxyl-3-nitrophenylacetyl-leucinyl-leucinyl-leucine-vinyl sulfone (NIP-L_3VS) (45), which inhibited the degradation of Phe-DHFR in egg extracts, also delayed apoptosis in the CSF-arrested extract by ~1 h (data not shown), similarly to the effect of Trp-Ala (Fig. 6B).

DISCUSSION

We report the following results:

(1) A whole-cell extract of *X. laevis* eggs contains the N-end rule pathway, as determined by monitoring the metabolic fates of purified, ^{35}S-labeled Ub-X-DHFR test proteins added to the extract. Among the cells of multicellular organisms, the N-end rule pathway has been identified and characterized in rabbit reticulocytes (18) and L cells, a mouse cell line (19, 26), but not, until now, in germ line cells.

(2) The partial N-end rule defined in egg extracts was found to be identical to the corresponding subsets of the N-end rules in rabbit reticulocytes and mouse fibroblasts. In particular, we showed that in egg extracts the destabilizing activity of N-terminal Asn, Asp, and Cys requires the presence of RNA (presumably tRNA), strongly suggesting that these are secondary (Asp and Cys) and tertiary (Asn) residues in the egg's N-end rule pathway, as they have been shown to be in other metazoan cells (see introduction for the terminology) (7).

(3) The activity of the N-end rule pathway did not change significantly upon shifts of the CSF-arrested egg extract to the cycling or the interphase-arrested state.

(4) Dipeptides bearing destabilizing N-terminal residues selectively and efficiently inhibited the degradation of N-end rule substrates in egg extract. In particular, dipeptides bearing the basic (type 1) primary destabilizing N-terminal residues inhibited the degradation of N-end rule substrates bearing type 1 but not type 2 N-terminal residues. A converse inhibition pattern was observed with dipeptides bearing bulky hydrophobic (type 2) N-terminal residues.

(5) The onset of apoptosis-like changes in egg extract (fragmentation of the added sperm nuclei and abrupt increase of the DEVD-specific protease activity (34, 43)) was found to be delayed by ~1 h in the presence of dipeptide Trp-Ala, which bears a type 2 destabilizing N-terminal residue, but not in the presence of dipeptides bearing type 1 destabilizing N-terminal residues.

The N-end rule pathway has been identified in all organisms examined, from mammals to fungi and bacteria, but the understanding of its functions remains incomplete (6, 7). The first clear example of a physiological function of the N-end rule pathway is its recently identified role in controlling the import of peptides in *S. cerevisiae* through the degradation of the transcriptional repressor Cup9p that down-regulates a peptide transporter (25, 46). It has also been reported that dipeptides bearing destabilizing N-terminal residues specifically inhibit differentiation of rat pheochromocytoma PC12 cells (47), the neurite outgrowth in amphibian neuroepithelial cells (48), and limb regeneration in the newt (49), suggesting a role for the N-end rule pathway in these biological processes. Unfortunately, none of these studies (47–49) included control experiments to verify that the added dipeptides actually inhibited degradation of a reporter N-end rule substrate in the target cells. Thus, it remains to be determined whether the observed biological effects of dipeptides (47–49) were caused by specific inhibition of the N-end rule pathway in the target cells or resulted from other effects of the added dipeptides.

Apart from the interest in determining whether the N-end rule pathway is present in germ line cells such as *Xenopus* eggs, one reason for initiating this work was the opportunity to test, through selective inhibition of the N-end rule pathway in egg extracts, whether this pathway plays a role in apoptosis. Several lines of evidence suggest a role for the Ub/proteasome system, of which the N-end rule pathway is a part, in regulating apoptosis (50–55). In particular, cell-penetrating proteasome inhibitors were reported to partially protect nonproliferating cells such as thymocytes and growth factor-deprived sympathetic neurons from apoptosis (52, 53). In contrast, the same proteasome inhibitors were reported to induce apoptosis in mitotically active cells (51, 54). An apoptosis-like process, induced by the stress of starvation and mediated by proteolysis, was also identified in *E. coli* (56). A key regulator of *E. coli* apoptosis is a short-lived protein MazE, which is degraded by ClpAP, a proteasome-like ATP-dependent protease that targets N-end rule substrates in *E. coli* (38). These and other considerations, including the possibility that cleavages by caspases can produce N-end rule substrates, led to the suggestion that the N-end rule pathway may function in the control of apoptosis (7).

Specific dipeptides, which have been shown to act as selective inhibitors of the N-end rule pathway in reticulocyte extracts (19, 27) and intact *S. cerevisiae* cells (28), were found to be largely ineffective with several lines of mammalian cells[3]. *Xenopus* egg extracts, in contrast to reticulocyte extracts, have been shown to undergo apoptosis-like changes (34, 36, 43), hence the choice of these extracts for the present study. We found that dipeptides bearing either type 1 (Arg, Lys) or type 3 (Ala) destabilizing N-terminal residues had no effect on the apoptotic changes in the extracts. In contrast, Trp-Ala, which bears a type 2 destabilizing N-terminal residue, inhibited apoptosis in a number of independent experiments with different preparations of egg extracts; in a minority of tests the effect was not observed. Tyr-His, a dipeptide bearing another type 2 destabilizing N-terminal residue, strongly inhibited the degradation of an N-end rule substrate such as Phe-DHFR in the extract, but was significantly less efficacious than Trp-Ala in its effect on the timing of apoptotic changes in the same extract. Taken together, the results of experiments with dipeptides as inhibitors of apoptosis indicate that some inhibitors of the N-end rule pathway, notably Trp-Ala, can cause a significant delay of apoptosis in the extract. A definitive test and further analysis of the thus suggested function of the N-end rule pathway in apoptosis will require mutants that eliminate this pathway in a multicellular organism without perturbing the rest of the Ub system. Construction of a mouse mutant that lacks the entire N-end rule pathway (through a deletion of the *Ubr1* gene that encodes N-recognin, the main recognition component of this proteolytic system) is under way[4].

In summary, we established the existence of the N-end rule pathway in germ line cells such as *Xenopus* eggs, showed that dipeptides are efficacious inhibitors of the N-end rule pathway in egg extracts, and examined the effects of dipeptides on apoptotic changes in these extracts.

ACKNOWLEDGMENTS

We are grateful to W. Dunphy for helpful discussions and the access to his laboratory's *Xenopus* facility, and to M. Bogyo and H. Ploegh for a gift of the NIP-L_3VS proteasome inhibitor. We thank E. Smirnova and members of the Varshavsky lab, especially L. Peck, G. Turner, and A. Webster, for their suggestions and for the comments on the manuscript. D.P. is a Fellow of the Leukemia Society of America. This work was supported by grants to A.V. from the National Institutes of Health (DK39520 and GM31530).

REFERENCES

1. Haas, A. J., and Siepman, T. J. (1997) *FASEB J.* **11,** 1257–1268.
2. Varshavsky, A. (1997) *Trends Biochem. Sci.* **22,** 383–387.
3. Hochstrasser, M. (1996) *Annu. Rev. Genet.* **30,** 405–439.
4. King, R. W., Deshaies, R. J., Peters, J. M., and Kirschner, M. W. (1996) *Science* **274,** 1652–1659.
5. Varshavsky, A. (1991) *Cell* **64,** 13–15.
6. Bachmair, A., Finley, D., and Varshavsky, A. (1986) *Science* **234,** 179–186.
7. Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 12142–12149.
8. Bachmair, A., and Varshavsky, A. (1989) *Cell* **56,** 1019–1032.
9. Johnson, E. S., Gonda, D. K., and Varshavsky, A. (1990) *Nature* **346,** 287–291.

---

[4] Y. T. Kwon and A. Varshavsky, unpublished data.

10. Hill, C. P., Johnston, N. L., and Cohen, R. E. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 4136–4140.
11. Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K., and Varshavsky, A. (1989) *Science* **243,** 1576–1583.
12. Pickart, C. M. (1997) *FASEB J.* **11,** 1055–1066.
13. Wilkinson, K. D. (1997) *FASEB J.* **11,** 1245–1256.
14. Hershko, A., and Ciechanover, A. (1992) *Annu. Rev. Biochem.* **61,** 761–807.
15. Finley, D., and Chau, V. (1991) *Annu. Rev. Cell Biol.* **7,** 25–69.
16. Coux, O., Tanaka, K., and Goldberg, A. L. (1996) *Annu. Rev. Biochem.* **65,** 801–817.
17. Hilt, W., and Wolf, D. H. (1996) *Trends. Biochem. Sci.* **21,** 96–102.
18. Lupas, A., Flanagan, J. M., Tamura, T., and Baumeister, W. (1997) *Trends Biochem. Sci.* **22,** 399–404.
19. Gonda, D. K., Bachmair, A., Wünning, I., Tobias, J. W., Lane, W. S., and Varshavsky, A. (1989) *J. Biol. Chem.* **264,** 16700–16712.
20. Balzi, E., Choder, M., Chen, W., Varshavsky, A., and Goffeau, A. (1990) *J. Biol. Chem.* **265,** 7464–7471.
21. Baker, R. T., and Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 12065–12074.
22. Bartel, B., Wünning, I., and Varshavsky, A. (1990) *EMBO J.* **9,** 3179–3189.
23. deGroot, R. J., Rümenapf, T., Kuhn, R. J., and Strauss, J. H. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 8967–8971.
24. Madura, K., and Varshavsky, A. (1994) *Science* **265,** 1454–1458.
25. Byrd, C., Turner, G. C., and Varshavsky, A. (1998) *EMBO J.* **17,** 269–277.
26. Lévy, F., Johnsson, N., Rümenapf, T., and Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 4907–4912.
27. Reiss, Y., Kaim, D., and Hershko, A. (1988) *J. Biol. Chem.* **263,** 2693–269.
28. Baker, R. T., and Varshavsky, A. (1991) *Proc. Natl. Acad. Sci. USA* **87,** 2374–2378.
29. Dunphy, W. G. (1994) *Trends Cell Biology* **4,** 202–207.
30. Johnston, J. A., Johnson, E. S., Waller, P. R. H., and Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 8172–8178.
31. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K. Eds. (1996) Current Protocols in Molecular Biology, Wiley–Interscience, New York.
32. Murray, A. W. (1991) *Methods Cell. Biol.* **36,** 581–605.
33. Grigoryev, S., Stewart, A. E., Kwon, Y. T., Arfin, S. M., Bradshaw, R. A., Jenkins, N. A., Copeland, N. J., and Varshavsky, A. (1996) *J. Biol. Chem.* **271,** 28521–28532.
34. Newmeyer, D. D., Farschon, D. M., and Reed, J. C. (1994) *Cell* **79,** 353–364.
35. Nicholson, D. W., Ali, A., Thornberry, N. A., Vaillancourt, J. P., Ding, C. K., Gallant, M., Gareau, Y., Griffin, P. R., Labelle, M., Lazebnik, Y. A., Munday, N. A., Raju, S. M., Smulson, M. E., Yamin, T.-T., Yu, V. L., and Miller, D. K. (1995) *Nature* **376,** 37–43.
36. Cosulich, S. C., Green, S., and Clarke, P. R. (1996) *Curr. Biol.* **6,** 353–364.
37. Johnson, E. S., Ma, P. C. M., Ota, I. M., and Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 17442–17456.
38. Tobias, J. W., Shrader, T. E., Rocap, G., and Varshavsky, A. (1991) *Science* **254,** 1374–1377.
39. Sagata, N. (1996) *Trends Cell Biol.* **6,** 22–28.
40. Furuno, N., Ogawa, Y., Iwashita, J., Nakajo, N., and Sagata, N. (1997) *EMBO J.* **16,** 3860–3865.
41. Evans, E. K., Lu, W., Strum, S. L., Mayer, B. J., and Kornbluth, S. (1997) *EMBO J.* **16,** 230–241.
42. Kluck, R. M., Bossy-Wetzel, E., Green, D. R., and Newmeyer, D. D. (1997) *Science* **275,** 1132–1136.
43. Kluck, R. M., Martin, S. J., Hoffman, B. M., Zhou, J. S., Green, D. R., and Newmeyer, D. D. (1997) *EMBO J.* **16,** 4639–4649.
44. Villa, P., Kaufmann, S. H., and Earnshaw, W. C. (1997) *Trends Biochem. Sci.* **22,** 388–393.
45. Bogyo, M., McMaster, J. S., Gaczynska, M., Tortorella, D., Goldberg, A. L., and Ploegh, H. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 6629–6634.
46. Alagramam, K., Naider, F., and Becker, J. M. (1995) *Mol. Microbiol.* **15,** 225–234.
47. Hondermarck, H., Sy, J., Bradshaw, R. A., and Arfin, S. M. (1992) *Biochem. Biophys. Res. Commun.* **30,** 280–288.
48. Maufroid, J. P., Bradshaw, R. A., Boilly, B., and Hondermarck, H. (1996) *Int. J. Dev. Biol.* **40,** 609–611.
49. Taban, C. H., Hondermarck, H., Bradshaw, R. A., and Boilly, B. (1996) *Experentia* **52,** 865–870.
50. Delic, J., Morange, M., and Magdelenat, H. (1993) *Mol. Cell. Biol.* **13,** 4875–4883.
51. Imajoh-Ohmi, S., Kawaguchi, T., Sugiyama, S., Tanaka, K., Omura, S., and Kikuchi, H. (1995) *Biochem. Biophys. Res. Commun.* **217,** 1070–1077.
52. Grimm, L. M., Goldberg, A. L., Poirier, G. G., Schwartz, L. M., and Osborne, B. A. (1996) *EMBO J.* **15,** 3835–3844.
53. Sadoul, R., Fernandez, P.-A., Quiquerez, A.-L., Martinou, I., Maki, M., Schröter, M., Becherer, J. D., Irmler, M., Tschopp, J., and Martinou, J.-C. (1996) *EMBO J.* **15,** 3845–3852.
54. Drexler, H. C. A. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 855–860.
55. Lopes, U. G., Erhardt, P., Yao, R., and Cooper, G. M. *J. Biol. Chem.* **272,** 12893–12896.
56. Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 6059–6063.

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 2094–2099, March 1998
Biochemistry

Codominant interference, antieffectors, and multitarget drugs

(pharmacology/cancer/codominance/selectivity)

ALEXANDER VARSHAVSKY[†]

Division of Biology, California Institute of Technology, Pasadena, CA 91125

Contributed by Alexander Varshavsky, December 19, 1997

ABSTRACT **The insufficient selectivity of drugs is a bane of present-day therapies. This problem is significant for antibacterial drugs, difficult for antivirals, and utterly unsolved for anticancer drugs, which remain ineffective against major cancers, and in addition cause severe side effects. The problem may be solved if a therapeutic agent could have a multitarget, combinatorial selectivity, killing, or otherwise modifying, a cell if and only if it contains a predetermined set of molecular targets and lacks another predetermined set of targets. An earlier design of multitarget drugs [Varshavsky, A. (1995) *Proc. Natl. Acad. Sci. USA* 92, 3663–3667] was confined to macromolecular reagents such as proteins, with the attendant difficulties of intracellular delivery and immunogenicity. I now propose a solution to the problem of drug selectivity that is applicable to small (≤1 kDa) drugs. Two ideas, codominant interference and antieffectors, should allow a therapeutic regimen to possess combinatorial selectivity, in which the number of positively and negatively sensed macromolecular targets can be two, three, or more. The nature of the effector and interference moieties in a multitarget drug determines its use: selective killing of cancer cells or, for example, the inhibition of a neurotransmitter-inactivating enzyme in a specific subset of the enzyme-containing cells. The *in vivo* effects of such drugs would be analogous to the outcomes of the Boolean operations "and," "or," and combinations thereof. I discuss the logic and applications of the antieffector and interference/codominance concepts, and the attendant problem of pharmacokinetics.**

The many successes of pharmacology (1, 2) do not include the problem of cancer. Major human cancers are incurable once they have metastasized. A few relatively rare cancers, such as testicular carcinoma in men, Wilms' kidney tumor, and some leukemias in children, can often be cured through chemotherapy but require cytotoxic treatments of a kind that cause severe side effects and are themselves carcinogenic (2).

The main reason for the failure of cytotoxic therapies is their insufficient selectivity for tumors. For example, treatments with radiation or alkylating agents perturb many functions that are common to all cells. The more selective cytotoxic drugs, for instance, methotrexate, taxol, and etoposide, perturb the functions of specific macromolecular targets (dihydrofolate reductase, microtubules, and topoisomerase II), but these targets are present in both normal and malignant cells (1, 2). Hence the low therapeutic index of anticancer drugs and their systemic toxicity at clinically relevant doses. Because the mitotic activity of cells in a tumor is often lower than the mitotic activity of normal cells in self-renewing tissues such as the bone marrow (3), one might not have expected these drugs to work at all—to have any preference for the killing of cancer cells. That such preference actually exists stems in part from the fact that tumor cells are often perturbed by their mutations into stress-hypersensitive states. Consequently, these tumor cells die an apoptotic death at the level of a drug-imposed metabolic stress that induces apoptosis in some but not in most of the organism's normal cells (3).

With some cancers, cytotoxic therapies are ineffective from the beginning. In other cases, these therapies yield a partial, sometimes clinically complete, but almost invariably transient remission of a cancer, in part because these treatments select for tumor cell variants that retain tumorigenicity but are more resistant to either apoptosis *per se* or a drug that induces apoptosis. Because a significant increase in the drug or radiation dosage is precluded by their low therapeutic index, these therapies become ineffective when resistant clones of malignant cells, selected by a drug treatment, present themselves as a cancer recurrence.

The failure of small cytotoxic drugs to produce a cure for cancer has given rise to other strategies, in particular the insightful suggestion that solid tumors can be targeted by selectively inhibiting neovascularization, a process that these tumors depend on for growing to a clinically significant size (4). Another approach, immunotoxins, involves the linking of a toxin to a ligand such as an antibody or a growth factor that binds to a target on the surface of tumor cells (5). Among the limitations of present-day immunotoxins is their incapacity, on entering a cell, to adjust their toxicity in response to the intracellular protein composition. Yet another approach is to enhance the ability of the immune system to identify and selectively destroy tumor cells. The current revival of this strategy holds the promise of a rational and curative treatment (6). Given the complicated regimens and unsolved problems of immunotherapies, it is clear that this and other recent approaches (7, 8) are motivated in part by the perception that small-drug pharmacology, so successful against bacterial infections, is unlikely to prove effective against cancer. In contrast to this view, the premise of the strategy described in the present work is that small anticancer drugs may become curative and free of severe side effects if a way is found to confer on these compounds a multitarget, combinatorial selectivity.

Most cancers are monoclonal: cell lineages of both the primary tumor and the metastases originate from a single founder cell. This cell is a breakthrough descendant of a cell lineage that has been accumulating mutations for some time, often in proximity to other neoplastic but still nonmetastatic cell lineages within an indolent proliferative lesion such as a benign tumor (9, 10). Given the monoclonality of a cancer, cells of both the primary tumor and the metastases share the initial mutations that yielded the founder cell, even if these cells differ at other loci that accumulated mutations in the course of the later tumor progression. Some of the early mutations are in genes that encode tumor suppressors (9, 11, 12). In most cancers, both alleles of a tumor suppressor gene are inactivated, sometimes through deletions that encompass the gene on the two homologous chromosomes. Thus, a monoclonal cancer, although heterogeneous genetically,

---

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/952094-6$2.00/0
PNAS is available online at http://www.pnas.org.

Abbreviation: IC, interference/codominance.

[†]To whom reprint requests should be addressed at: Division of Biology, 147-75, Caltech, 1200 East California Boulevard, Pasadena, CA 91125. e-mail: avarsh@cco.caltech.edu.

always contains a set of founder mutations that is shared by all of its cells.

A drug that kills a cell only if it lacks a specific macromolecular target would distinguish tumor cells from many other cells of an organism, provided that the target is a product of a gene that had been deleted or inactivated in this cancer at the stage of its founder cell. Such a drug may be especially selective against cancers that lack a gene for a ubiquitously expressed tumor suppressor, for example, the retinoblastoma (Rb) protein (11, 12). An example of the negative-target approach is the use of a mutant adenovirus that replicates selectively in human cancer cells lacking the tumor suppressor p53 and has been shown to kill these cells in a model setting (8).

However, other tumor suppressors may not be expressed at comparable levels in most cells. A drug that kills a cell if it lacks a nonubiquitous tumor suppressor would be toxic to a subset of normal cells as well. This problem could be reduced through the use of a drug that is toxic only if a cell lacks two specific macromolecules, termed negative targets. Two judiciously chosen negative targets may, together, suffice to distinguish all of the cancer cells from all of the organism's normal cells. If they do not, a third negative target that had been deleted or rendered defective in a given cancer can be employed as well. This strategy requires a drug that possesses the ability to kill a cell if it lacks two or more of the predetermined targets, but would spare a cell containing either one of these targets.

Other changes in a founder cell may involve a missense mutation, an amplification and overexpression, an ectopic expression, or a translocation/fusion of a specific protooncogene such as, for example, *Ras* or *Myc* (9, 10, 13). A single oncoprotein may not be a unique enough target by itself, for reasons similar to those described above in the context of negative targets. However, a combination of two or more distinct oncoproteins that were either mutated or inappropriately expressed in the founder cell can be employed to formulate the unique multiprotein signature of a specific cancer that comprises both positive and negative targets.

These considerations suggest that a conditionally cytotoxic therapeutic regimen that is exquisitely specific for a given cancer, and therefore would eliminate it without significant side effects, must possess, in most cases, a multitarget, combinatorial (positive/negative) selectivity of the kind defined above. Conversely, even an informed choice of the molecular target for a single-target drug may not suffice to define unambiguously the cell type to be eliminated. Note that simply combining two single-target drugs against two different targets in a multidrug regimen would not yield a multitarget selectivity, because the two drugs together would perturb not only cells containing both targets but also cells containing either one of the targets.

Although the problem of insufficient selectivity is not as acute with noncytotoxic drugs, it is relevant to them as well. Among the multitude of examples are side effects of therapies with antipsychotic agents. The side effects are caused in part by the insufficient molecular specificity of drugs, which is exemplified by the ability of antidepressants that inhibit monoamine oxidase to perturb other proteins as well (2). This difficulty will continue to abate with the development of more specific single-target inhibitors. But an entirely distinct, major, and unsolved problem with inhibitors as drugs is the current impossibility of restricting their action to a specific subset of cells among those that contain the inhibitor's target. For example, even an exquisitely specific inhibitor of a clinically relevant enzyme is likely to have side effects, because the target enzyme is present, in most cases, not only in the cells where its inhibition is clinically beneficial but also in the cells where its inhibition is physiologically inappropriate. The present work describes a possible solution of this problem.

A previously proposed approach to designing multitarget drugs utilized degradation signals (degrons) and analogous signals that exhibit the property of codominance (14, 15). As a result, this strategy was confined to macromolecular reagents such as proteins, with the attendant problems of immunogenicity, extravasation, and intracellular delivery. The latter difficulty is especially significant, because either gene-therapy or direct-delivery methods for introducing large molecules into cells work reasonably well with cells in culture but are still inefficient with cells in an intact organism. The challenge, then, is to attain a multitarget, combinatorial selectivity in the setting of small (≤1 kDa) drugs, where the immunogenicity and delivery problems are less severe. A solution, described below, invokes a modification of the earlier idea of codominant interference (14) in conjunction with the new concept of antieffectors. This solution is applicable to either cytotoxic or noncytotoxic therapies.

RESULTS AND DISCUSSION

Multitarget Compounds Specific for Negative Targets: The Concept of Codominant Interference. Previous work (14) suggested that the property of codominance, characteristic of degradation signals (degrons) and many other signals in biopolymers, can be employed to design protein-based reagents that possess multitarget, combinatorial selectivity of the kind defined above. Codominance refers to the ability of two or more signals in the same molecule to function independently and not to interfere with each other. It is shown below that a distinct version of the interference/codominance (IC) concept (14) is applicable to small (≤1 kDa) compounds. Consider a reagent containing three small moieties **a**, **b**, and **i**, which can bind, respectively, to three macromolecular targets **A**, **B**, and **I**. Because the moieties **a**, **b**, and **i** are much smaller than the macromolecules **A**, **B**, and **I**, it should be possible to arrange these moieties in the compound **abi** in such a way that the binding of **A** or **B** to **a** or **b** would preclude, through steric hindrance, the binding of moiety **i** to **I** (Fig. 1).

That the interactions of a small bipartite compound with its two macromolecular ligands can be made mutually exclusive is expected from basic physicochemical considerations. This has also been demonstrated directly, in a context unrelated to the present discussion. When lisinopril, an inhibitor of the angiotensin-converting enzyme (ACE), was connected, via an 11-atom linker, to the biotin moiety, the resulting bivalent compound could bind to and inhibit ACE in the absence but not in the presence of the biotin-binding protein streptavidin (16). Small compounds comprising two linker-connected moieties such as cyclosporin and FK506, which are specific for two macromolecular targets, have previously been employed as *in vivo* dimerization devices, making it possible to bring together two otherwise noninteracting proteins (17). However, the linker moiety of these bipartite compounds was chosen to allow simultaneous interactions with the targets, in contrast to the mutual exclusivity of interactions in the IC approach (Fig. 1).

If the moiety **i** is an inhibitor of an essential cellular enzyme **I**, the presence of the macromolecular targets **A** or **B** in a cell would reduce the inhibition of enzyme **I** by **abi**, because the complexes **abi-A** and **abi-B** would be mutually exclusive with the complex **abi-I** (Fig. 1). Note that **A** and **B** are codominant in their ability to reduce the inhibition of **I** by **abi**. Therefore, there is, formally, no limit on the number of **a**, **b**-like competition modules that can be used to construct an **abi**-like compound whose activity is sensitive to the presence of several distinct macromolecules, called negative targets. The fractional occupancy of the macromolecular targets **A**, **B**, and **I** by the **a**, **b**, and **i** moieties of **abi** would be determined in part by the targets' intracellular concentrations. There are also specific pharmacokinetic constraints on the selectivity of **abi**, an issue discussed below.

A tabulation of the relative toxicities of **abi** for cells that either lack or contain the negative targets **A** and **B** is shown in Fig. 2. It can be seen that **abi** would be relatively nontoxic to three of the four cell types and toxic exclusively to the cells that lack both **A** and **B** (Fig. 2). Thus, the IC concept allows the construction of

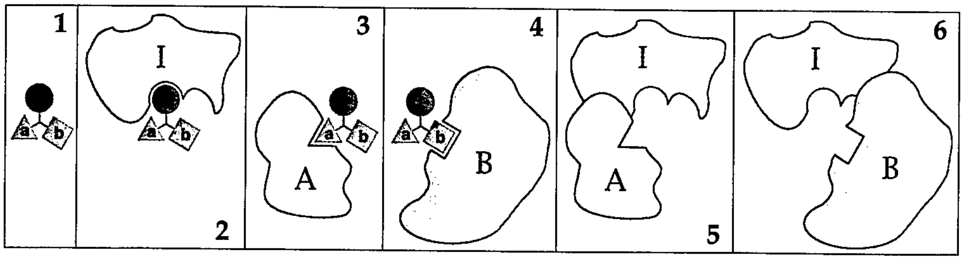


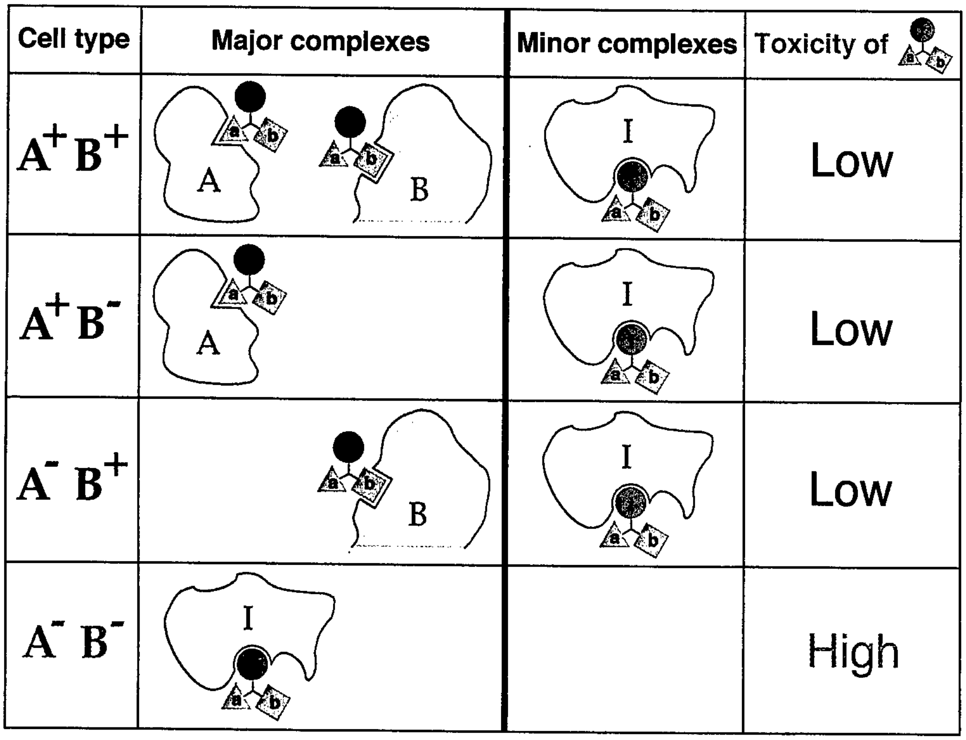
FIG. 1. The interference/codominance concept. (*1–4*) A small (<1 kDa) moiety **i** is linked to two other small moieties, **a** and **b**. The moieties **a**, **b**, and **i** are ligands of the macromolecules **A**, **B**, and **I**, respectively. The distances between **a**, **b**, and **i**, and their mutual arrangement in the tripartite compound **abi** are such that the interaction **i-I** is mutually exclusive with either the interaction **a-A** or the interaction **b-B**. Specifically, the macromolecule **I** in its complex with the small moiety **i** would sterically clash with the macromolecules **A** or **B** if either **A** or **B** is positioned to bind **a** or **b** of **abi** (*5* and *6*). In the diagram, the interactions **a-A** and **b-B** are also mutually exclusive, but this constraint is not essential. (Note that if the interactions **a-A** and **b-B** were mutually nonexclusive, the compound **abi** would promote the binding of **A** to **B**.) The codominance aspect of the IC concept allows this design to accommodate more than two of the **a**, **b**-like competition modules (not shown). In Figs. 2–4, the **i** moiety is an inhibitor of an essential enzyme **I**. In fact, the only constraint on the identities of **i** and **I** is the requirement for an **i-I** interaction to alter the functional activity of a macromolecule **I**. In other words, the choice of **I** is determined by the intended effect of the (unsequestered) compound **abi** (see the main text).

small compounds that exhibit multitarget selectivity for negative targets. One more idea is required to accomplish the same for positive targets and to link the two strategies.

Multitarget Compounds Specific for Positive Targets: The Concept of Antieffectors. Consider a small compound **i*** that binds to enzyme **I** in the vicinity of its active site, but does not perturb the catalytic activity of **I** toward its physiological substrates (Fig. 3*B*). Suppose further that the compound **i***, termed an antiinhibitor, was designed to interfere, sterically, with the binding of an inhibitor **i** to the enzyme's active site while at the same time allowing the binding of physiological substrates. One way to achieve this would be to endow either **i**, or **i***, or both of them with a set of chemical groups, termed a "bump," whose function is to produce steric hindrance that makes the interactions **i-I** and **i*-I** mutually exclusive (Fig. 3 *A* and *B*). A moiety that functions as a bump may also be designed to enhance specific binding of either the inhibitor **i** or the antiinhibitor **i*** to enzyme **I**, but this consideration is secondary to the bump's essential purpose.

In one application of the antiinhibitor **i***, it is linked to **c**, a small moiety that can bind to a macromolecular target **C**. The mutual arrangement of **i*** and **c** in **ci*** is such that the interactions of **ci*** with **I** and **C** are mutually exclusive. In the absence of **C**, **ci*** would compete with **abi** for the binding to enzyme **I**, thereby partially protecting **I** from inhibition by **abi** (Fig. 3*C*). This protective effect of **ci*** would be suppressed in the presence of its macromolecular target **C** (Fig. 3*D*). In the logic of codominance, discussed above in the context of negative targets, a compound bearing an antiinhibitor moiety **i*** could contain more than one **c**-like moiety. For example, a compound **cdi***, whose moieties **c** and **d** can bind, respectively, to the macromolecules **C** and **D**, would reduce the inhibition of enzyme **I** by **abi** only in the absence of both **C** and **D**. Yet another pattern of multitarget selectivity can be produced, in this context, by separating the competition moieties **c** and **d**. The resulting **ci*** and **di***, if administered together with **abi**, would reduce the inhibition of enzyme **I** by **abi** if just one of the targets, **C** or **D**, is absent. As shown below, the key merit of the antiinhibitor idea is that it allows the effect of a single inhibitor **i** to be modulated by both negative and positive macromolecular targets.

On the Difference Between Antieffectors and Antagonists. The distinctions between substrates, inhibitors, and antiinhibitors were described above. The concept of antieffectors is also relevant to ligand-binding biopolymers other than enzymes. For example, an agonist binds to its receptor and evokes a physiological response. An antagonist binds to a site of the receptor that overlaps with the agonist-binding site, does not



Cell type	Major complexes	Minor complexes	Toxicity of
$A^+ B^+$	a b A a b B	I a b	Low
$A^+ B^-$	a b A	I a b	Low
$A^- B^+$	a b B	I a b	Low
$A^- B^-$	I a b		High

FIG. 2. Multitarget selectivity of a compound that utilizes interference/codominance. This diagram tabulates the relative toxicities of the compound **abi** for cells that either lack or contain macromolecular targets **A** and **B**. The **i** moiety of the compound **abi** (see the legend to Fig. 1) inhibits an essential enzyme **I**. The interaction **i-I** is mutually exclusive with the interaction **a-A** and the interaction **b-B**, the macromolecules **A** and **B** being negative targets of **abi**. It is assumed that concentrations of the targets **A** and **B** in cells that contain at least one of them significantly exceed the concentration of **I** (see the main text). In $A^+ B^+$, $A^+ B^-$, and $A^- B^+$ cells, the enzyme **I** would be at most partially inhibited by the **i** moiety of **abi**, because of the competing interactions of **abi** with **A** and/or **B**. By contrast, in $A^- B^-$ cells, the bulk of **abi** molecules would be available for interaction with **I**, resulting in the selective toxicity of **abi** to these cells. The selectivity pattern of **abi** requires that certain pharmacokinetic conditions are met as well (see the main text). Note that the physiological effects and the uses of **abi**-type compounds are not confined to cytotoxic regimens.

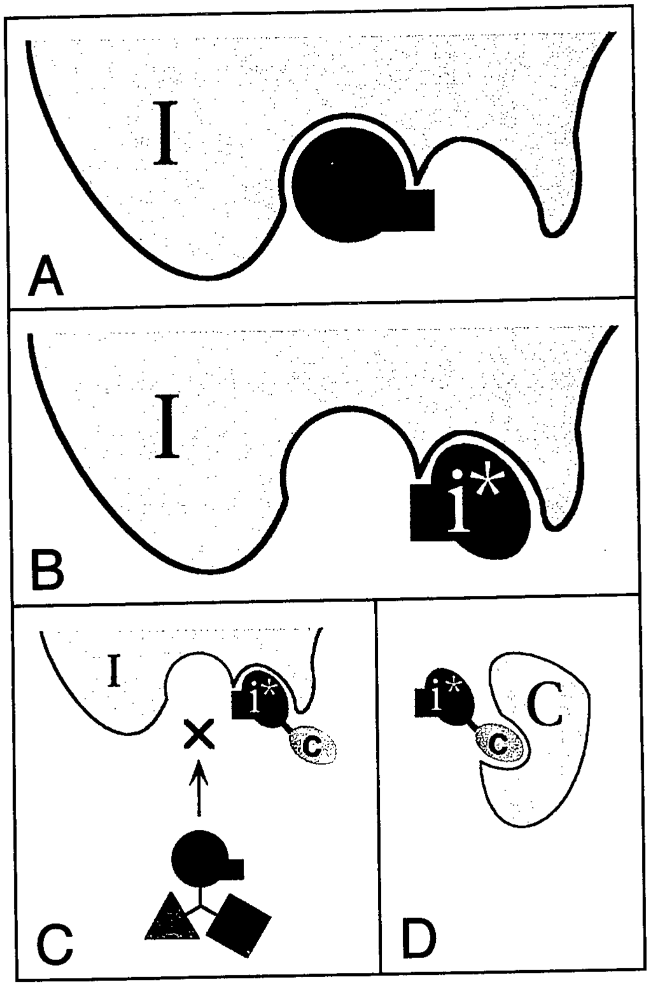


FIG. 3. The antieffector concept. The particular case illustrated here is that of an antiinhibitor **i***, defined as a compound whose binding to an enzyme **I** does not inhibit the activity of **I** but does preclude the inhibition of **I** by an inhibitor **i**. In this example, the antiinhibitor **i*** has the following properties. First, it binds to **I** in the vicinity of the I's active site, but does not perturb the catalytic activity of **I** toward its physiological substrates. Second, **i***, in its bound state, sterically interferes with the interaction between **I** and its inhibitor **i**. To implement the second condition, either **i**, or **i***, or both of them bear additional moiety, a "bump," denoted by the rectangular protrusions in **i** and **i***. The function of the bump is to produce steric hindrance that makes the interactions **i-I** and **i*-I** mutually exclusive. The inhibitor **i** described here and in the main text is a competitive inhibitor, but **i** could be a noncompetitive inhibitor as well. An allosteric antiinhibitor, which functions through binding to a remote site of enzyme **I**, is yet another possibility. (*A*) A complex of the enzyme **I** with its inhibitor **i**. (*B*) A complex of **I** with its antiinhibitor **i***. Note that the bumps of the bound **i** and **i*** spatially overlap. (*C*) The antiinhibitor **i*** is linked to **c**, a small moiety that can bind to a macromolecular target **C**. The design of **ci*** is analogous to **abi** (Figs. 1 and 2), in that the interactions of **ci*** with **I** and **C** are mutually exclusive. When **ci*** is bound to **I**, the inhibitor **i**, shown here as a part of the compound **abi** (Figs. 1 and 2), is unable to bind to and inhibit the enzyme **I**. (*D*) A complex between **ci*** and its macromolecular target **C**. This complex, being mutually exclusive with the **ci*-I** complex, reduces the ability of **ci*** to protect the enzyme **I** from inhibition by **abi**.

activate the receptor, and in addition precludes the binding of agonist (1, 2). By contrast, an antieffector, which would be called, in this setting, an antiantagonist, binds to the receptor in such a way that the receptor can still bind, and respond to, the agonist, but cannot bind the antagonist. To this end, either an antagonist, or an antiantagonist, or both must possess a bump, an additional moiety described above in the context of enzymes and antiinhibitors (Fig. 3). The idea of antieffectors is thus distinct from that of antagonists or inhibitors and is new, to the best of my knowledge. I am also not aware of a naturally occurring pair of compounds that satisfy the definition of effectors/antieffectors in a physiologically relevant setting.

Interference/Codominance and Antieffector in a Regimen That Possesses Combinatorial Selectivity. Applying the IC and antieffector concepts together yields regimens that possess true combinatorial selectivity, i.e., sensitivity to both negative and positive targets. Consider a population of cells that either contain or lack the macromolecular targets **A**, **B**, and **C**. Our aim is to devise a treatment that would be toxic to cells that lack **A** and **B** but contain **C** (A^- B^- C^+ cells) and relatively nontoxic to the other cell types (Fig. 4). A regimen of two compounds, **abi** (Fig. 3*C*) and **ci*** (Fig. 3*D*), has the requisite selectivity, as shown in Fig. 4, which tabulates the outcomes of this treatment for different cell types. Specifically, in the absence of **C** (four cell types out of eight), the antiinhibitor-containing **ci*** would compete with the inhibitor-containing **abi** for binding to the essential enzyme **I**, thereby reducing the inhibition of **I** by **abi**, and hence reducing the toxicity of **abi**. In three other cell types, whose common property is the presence of **C** and at least one of the other two targets, **A** or **B**, the antiinhibitor-containing **ci*** would be largely sequestered by **C**, and hence inactive, but the inhibitor-containing **abi** would be sequestered as well, by either **A** or **B**. In only one type of cells, those that lack **A** and **B** but contain **C** (A^- B^- C^+ cells), is the inhibitor **abi** fully available for interaction with **I**, resulting in higher toxicity (Fig. 4). The differences in the toxicity of **abi** to different cell types would be determined by the relative stoichiometries and absolute concentrations of the cellular targets involved (**A**, **B**, **C**, and **I**), by the affinities of the moieties **a**, **b**, **c**, **i**, and **i*** for these targets, and by the pharmacokinetic properties of **abi** and **ci***.

Straightforward variations of the **abi** and **ci*** designs that utilize IC and the properties of antiinhibitors would allow selective targeting of any one of the eight cell types that differ by the presence or absence of three macromolecular targets. Moreover, there is no formal limit on the total number of negative and/or positive targets that can be simultaneously sensed by regimens that employ **abi**- and **ci***-type compounds bearing multiple interference moieties. Note that the cell type selectivity of regimens such as **abi** + **ci*** (Fig. 4) is analogous to the outcomes of the Boolean operations "and," "or," and combinations thereof.

Stoichiometries, Affinities, and Pharmacokinetics. The selectivity of the proposed compounds results from mutually exclusive, competing interactions between individual moieties of these compounds and their macromolecular targets (Figs. 2–4); hence, the importance of the targets' intracellular concentrations, relative to each other and the enzyme **I**, which is inhibited by an effector moiety of these drugs. The choice of **I** is not confined to essential enzymes. The target **I** could be, for instance, a DNA-binding repressor of terminal differentiation, a repressor of apoptosis, or, in the example of a noncytotoxic therapy, a neurotransmitter-inactivating enzyme. In other words, the choice of **I** is determined by the intended effect of the (unsequestered) compound **abi**.

The sequestration of **abi**- and **ci***-type compounds by their macromolecular ligands **A**, **B**, and **C** serves to prevent their binding to the enzyme **I** (Figs. 3 and 4). Therefore, in schemes of the type considered above, the molar concentration of **I** should be significantly (if possible, considerably) lower than the molar concentrations of **A**, **B**, and **C**. In addition, the concentration of **ci*** in a **ci*** + **abi** regimen should significantly exceed that of **abi**, because **ci*** is the sole obstacle to the inhibition of enzyme **I** by **abi** in the A^- B^- C^- cells (Fig. 4). It is assumed, furthermore, that the total intracellular concentrations of **abi** and **ci***, bound and unbound, would remain significantly below the concentrations of the interference targets **A**, **B**, and **C**. The affinities of **A**, **B**, **C**, and **I** for the

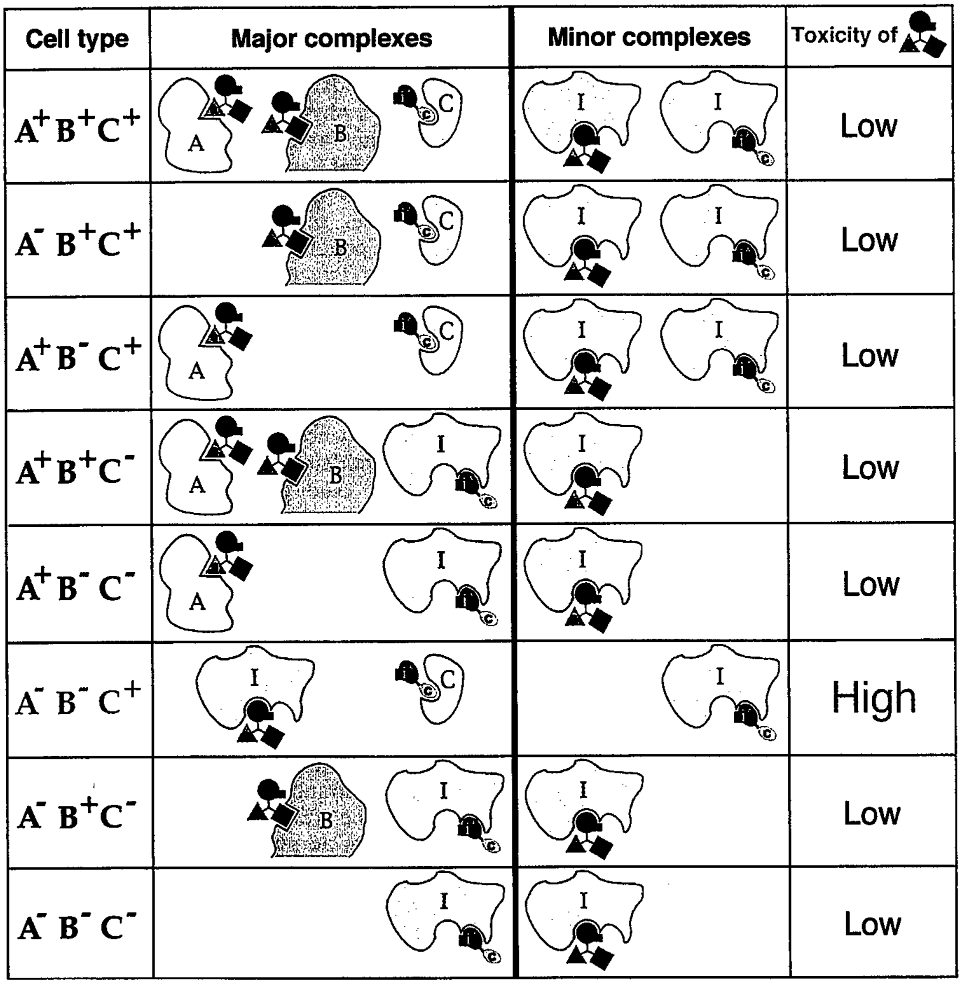


FIG. 4. Combinatorial (positive/negative) selectivity of a regimen that utilizes interference/codominance and antiinhibitor. The diagram tabulates the relative toxicities of the compound **abi** in the setting of a two-compound treatment of cells that either lack or contain the macromolecular targets **A**, **B**, and **C**. The inhibitor-containing compound **abi** is described in the main text and Figs. 1, 2, and 3*C*. The antiinhibitor-containing compound **ci*** is described in the main text and Fig. 3. **A** and **B** are negative targets, in that they reduce, through the binding to the moieties **a** and **b** of **abi**, the inhibition of an essential enzyme **I** by **abi**. **C** is a positive target, in that it reduces, through the binding to the **c** moiety of **ci***, the binding of **ci*** to enzyme **I**. This results in a larger fraction of enzyme **I** available for the inhibition by **abi**. It is assumed that the concentrations of **A**, **B**, and **C** in cells that contain them significantly exceed the concentration of **I** (see the main text). In all of the cell types except $A^- B^- C^+$, the enzyme **I** would be at most partially inhibited by the **i** moiety of **abi**, because of the competing interactions of **abi** with **A** and/or **B**, and also because in C^- cells a fraction of enzyme **I** would be protected from the inhibition by **abi** through the interaction of **I** with the antiinhibitor moiety **i*** of **ci***. By contrast, in $A^- B^- C^+$ cells, the antiinhibitor-containing **ci*** would be sequestered by **C**, whereas **abi** would not be sequestered by **A** or **B**, which are absent from these cells. As a result, a larger fraction of the inhibitor-containing **abi** molecules would be available for the interaction with **I**, resulting in the selective toxicity of **abi** to $A^- B^- C^+$ cells. The selectivity pattern of **abi** requires that certain pharmacokinetic conditions are met as well (see the main text). Note that the physiological effects and the uses of **abi**-type compounds are not confined to cytotoxic regimens.

respective moieties of the compounds **abi** and **ci***, and also the targets' intracellular locations are among the independent parameters that can be varied in designing these compounds.

Yet another, and major, constraint on the nature and pharmacokinetics of multitarget drugs stems from the fact that the selectivity patterns described above (Figs. 2 and 4) may not be observed under equilibrium conditions, where the influx of a drug into cells equals its outflux. To illustrate one clear difficulty, let us oversimplify and suppose that an **abi**-type compound is metabolically inert, in addition to being capable of crossing the plasma membranes and other lipid bilayers. If **abi** is initially outside the cells, and if the extracellular pool of **abi** (e.g., in the blood plasma) is large enough, it can be shown that the subsequently reached equilibrium state would be characterized by equal concentrations of the free **abi** in both the $A^+ B^+$ and $A^- B^-$ cells, thereby resulting in the equal occupancies of the enzyme **I** by **abi** in these cells, contrary to the pattern illustrated in Fig. 2.

By contrast, the selectivity patterns of Figs. 2 and 4 would be observed during the initial influx of drugs into cells. Thus, one requirement for the multitarget selectivity of **abi**- and **ci***-type regimens is the avoidance of equilibrium states such as the one described above. This and related considerations indicate that despite the logical simplicity of the proposed designs, their implementation will have to address pharmacokinetic problems that do not necessarily arise with single-target drugs.

Selection of Targets and Construction of Multitarget Drugs. Appropriate macromolecular targets of the **A**–**C** class (Figs. 3 and 4) are suggested by the protein composition of the tumor cells to be eliminated. The choice of an essential intracellular enzyme **I** (Figs. 2–4) is determined by the presence of **I** at least in tumor cells, its physiological concentration, and the feasibility of an efficacious inhibitor of **I**. Among potentially suitable enzymes for which cell-penetrating inhibitors already exist is dihydrofolate reductase. Its high-affinity inhibitors include methotrexate, which enters cells through carrier-mediated pathways, and the more lipophilic trimetrexate, which can enter cells by diffusing through lipid bilayers (18).

Each of the modules in the **abi**- and **ci***-type compounds (Figs. 1–4) would bind its macromolecular target in the absence of the other modules. Therefore, the cytotoxic **i**-type modules of IC-based compounds would be similar to the stand-alone cytotoxic drugs of today. By contrast, the interference modules of these compounds, i.e., their **a**-, **b**-, and **c**-type moieties (Figs. 1–3), are supposed to bind to their macromolecular targets but preferably not impair them functionally. This specification of a competition module simplifies its design in comparison to that of inhibitors, because many sites on the target's surface, and not just the active site, would be acceptable.

Antiinhibitors (Fig. 3) are a new class of physiologically active compounds. The opportunities and problems of their design are similar to those for the interference moieties **a**–**c** (Figs. 1–3), but there are two other difficulties as well. First, the antiinhibitor moiety **i*** must bind in the vicinity of, but not at, the active site of enzyme **I**. Second, the moiety **i*** must also bear chemical groups (a bump) whose function is to preclude, through steric hindrance, the binding of the inhibitor moiety **i** to the enzyme **I** (Fig. 3).

A substrate-binding cleft is not the only indentation in a folded protein molecule. The other clefts tend to be smaller, but they are present as well (19), and some of them may be located next to the enzyme's active site (Fig. 3). In addition, even relatively flat molecular surfaces can be, in principle, the sites of high-affinity interactions with small ligands (20). These optimistic comments notwithstanding, the development of antiinhibitors is certain to be a complex undertaking. As discussed above, the pharmacokinetic aspects of the proposed designs are also complex. Yet simplicity is good only if it works. Single-target anticancer drugs remain unsatisfactory, in spite of decades of immense effort. It may therefore be wise to attempt a more complex but also more effective solution.

A recent advance in drug design, termed SAR by NMR (structure–activity relationships by nuclear magnetic resonance), provides an especially promising route to constructing ligands for specific regions of a protein molecule (21). In this approach, a library of small molecules is screened for binding to an ^{15}N-labeled protein by using NMR, which can detect weak interactions, and in addition assigns them to specific nitrogens of a protein, thereby identifying the site of binding. Finding two small compounds that bind to adjacent patches of the target protein molecule and covalently linking these compounds produces a higher-affinity ligand. This powerful strategy (21), which already yielded tightly binding ligands of specific proteins, may prove sufficient for constructing the **a–c** competition modules and the **i*** antiinhibitor modules of the proposed designs (Figs. 1–4).

Noncytotoxic Multitarget Drugs. Many useful drugs are the inhibitors of intracellular enzymes that are not essential for cell viability (2). The problem of insufficient selectivity is relevant to these drugs as well. For example, even an exquisitely specific inhibitor of a clinically relevant enzyme is likely to have significant side effects, because the target enzyme is present, in most cases, not only in the cells where its inhibition is clinically beneficial but also in the cells where its inhibition is physiologically inappropriate. The logic of **abi**-type inhibitors (Figs. 1 and 2) and **ci***-type antiinhibitors (Figs. 3 and 4) is applicable in these settings, because an informed choice of the competition moieties **a**, **b**, and **c** would sharpen up the cell selectivity of the inhibitor moiety **i** in the way described above for cytotoxic drugs (Fig. 4), resulting in the inhibition of the (nonessential) enzyme **I** in a predetermined subset of the enzyme-containing cells. Note that the same considerations apply to extracellular settings as well. The examples above are but a glimpse of the drug-engineering vistas that are opened up by the IC and antieffector concepts. At the same time, there are significant pharmacokinetic constraints on the properties of the proposed drugs, as discussed above. These constraints are likely to complicate the implementation of the IC/antieffector strategies.

The Problem of Drug Resistance. With small anticancer drugs that are in use today, the macromolecular target of a drug serves two distinct functions. First, the target is a cell-selectivity determinant that may bias the treatment against tumor cells. Second, the target is also a device whose inhibition by the drug brings about the desired effect, e.g., cell death. Consequently, when drug-resistant tumor cells, selected by a drug treatment, present themselves as a cancer recurrence, the necessity of employing another therapeutic agent (if such an option exists) robs the physician of whatever cell-selectivity advantage there was with the earlier drug.

The situation is qualitatively different with IC-based compounds. Suppose that a treatment that included the drug **abi** (Figs. 1 and 2) results in the appearance of **abi**-resistant tumor cells that contain, for example, an altered or overproduced enzyme **I**. If so, replacement of the **i** moiety by another small cytotoxic moiety, specific for another essential enzyme, would retain the cell selectivity of the new **ab**-containing drug. Thus, one advantage of modularity inherent in the designs of IC/antiinhibitor-based compounds (Figs. 1–4) lies in the separation of the effector aspect of a drug from its selectivity aspect. As a result, once an efficacious arrangement of the selectivity modules in **abi**- or **ci***-type compounds has been identified, it can be reutilized in drugs bearing effector moieties other than **i** and **i***.

Concluding Remarks. The above considerations are based on the existing understanding of single-target drugs and on the notion of steric hindrance. By introducing the new concept of antieffectors and a modification of the previously proposed idea of codominant interference (14), we can now attempt the construction of small modular compounds that possess a multitarget, combinatorial selectivity (Fig. 4). The IC/antieffector strategies are not confined to cytotoxic therapies and are relevant, in principle, to all pharmacological settings. As indicated above, one expected difficulty in implementing these strategies stems from significant pharmacokinetic constraints that do not necessarily arise with single-target drugs.

This work was motivated by the premise that the confinement of anticancer drug research and development to single-target compounds will prove insufficient for the task at hand, because even the informed choices of targets for such drugs may not define unambiguously enough the cell type to be eliminated. The remedy, described above, is to aim for drugs that possess qualitatively different selectivity—multitarget and combinatorial. If this view is correct, the future ascent of multitarget drugs may transform not only the treatment of cancer but also approaches in other settings where the killing or modification of undesirable cells or organelles is carried out in the presence of nearly identical cells or organelles that must be spared. These applications of multitarget drugs encompass more discriminating antiviral and antifungal therapies, as well as the selective killing of activated lymphocytes in autoimmune diseases and the selective elimination of damaged mitochondria in aging cells (14, 15). In yet another class of applications, a noncytotoxic multitarget drug would be used to inhibit a clinically relevant nonessential enzyme in a specific subset of the enzyme-containing cells, thereby retaining the benefits of inhibition while reducing its side effects.

I thank L. Peck, G. Turner, D. Anderson, A. Rich, S. Mayo, E. Berezutskaya, and A. Kashina for comments on the manuscript. Studies in my laboratory are supported by grants from the National Institutes of Health.

1. Gilman, A. G., Rall, T. W., Nies, A. S. & Taylor, P. (1990) *The Pharmacological Basis of Therapeutics* (Pergamon, New York).
2. Munson, P. L., Mueller, R. A. & Breese, G. R. (1996) *Principles of Pharmacology* (Chapman & Hall, New York).
3. Waldman, T., Zhang, Y., Dillehay, L., Yu, J. K., Vogelstein, B. & Williams, J. (1997) *Nat. Med.* **3**, 1034–1036.
4. Folkman, J. (1985) *Adv. Cancer Res.* **43**, 175–203.
5. Thrush, G. R., Lark, L. R., Clinchy, B. C. & Vitetta, E. S. (1996) *Annu. Rev. Immunol.* **14**, 49–71.
6. Rosenberg, S. A. (1996) *Annu. Rev. Med.* **47**, 481–491.
7. da Costa, L. T., Jen, J., He, T.-C., Chan, T. A., Kinzler, K. W. & Vogelstein, B. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 4192–4196.
8. Bischoff, J. R., Kirn, D. H., Williams, A., Heise, C., Horn, S., Muna, M., Ng, L., Nye, J. A., Sampson-Johannes, A., Fattaey, A. & McCormick, F. (1996) *Science* **274**, 373–376.
9. Bishop, J. M. (1995) *Genes Dev.* **9**, 1309–1315.
10. Kinzler, K. W. & Vogelstein, B. (1996) *Cell* **87**, 159–170.
11. Weinberg, R. A. (1995) *Cell* **81**, 323–330.
12. Knudson, A. G. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 10914–10921.
13. Hunter, T. (1997) *Cell* **88**, 333–346.
14. Varshavsky, A. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 3663–3667.
15. Varshavsky, A. (1996) *Cold Spring Harbor Symp. Quant. Biol.* **60**, 461–478.
16. Bernstein, K. E., Welsh, S. L. & Inman, J. K. (1990) *Biochem. Biophys. Res. Commun.* **167**, 310–316.
17. Crabtree, G. R. & Schreiber, S. L. (1996) *Trends Biochem. Sci.* **21**, 418–422.
18. Takimoto, C. H. & Allegra, C. J. (1995) *Oncology* **9**, 649–659.
19. Laskowski, R. A., Luscombe, N. M., Swindells, M. B. & Thornton, J. M. (1996) *Protein Sci.* **5**, 2438–2452.
20. Mattos, C. & Ringe, D. (1996) *Nat. Biotech.* **14**, 595–599.
21. Hajduk, P. J., Meadows, R. P. & Fesik, S. W. (1997) *Science* **278**, 497–499.

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 7898–7903, July 1998
Biochemistry

The mouse and human genes encoding the recognition component of the N-end rule pathway

(ubiquitin/proteolysis/E3/N-recognin/*Ubr1*)

YONG TAE KWON[*], YUVAL REISS[†], VICTOR A. FRIED[‡], AVRAM HERSHKO[§], JEONG KYO YOON[*], DAVID K. GONDA[¶], PITCHAI SANGAN[¶], NEAL G. COPELAND[||], NANCY A. JENKINS[||], AND ALEXANDER VARSHAVSKY[*,**]

[*]Division of Biology, California Institute of Technology, Pasadena, CA 91125; [†]Department of Biochemistry, Tel Aviv University, Tel Aviv 69978, Israel; [‡]Department of Cell Biology and Anatomy, New York Medical College, Valhalla, NY 10595; [§]Unit of Biochemistry, Faculty of Medicine, Technion, Haifa 31096, Israel; [¶]Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT 06520-8024; and [||]Mammalian Genetics Laboratory, Advanced BioScience Laboratories-Basic Research Program, National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, MD 21702

Contributed by Alexander Varshavsky, May 5, 1998

ABSTRACT The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. The N-end rule pathway is one proteolytic pathway of the ubiquitin system. The recognition component of this pathway, called N-recognin or E3, binds to a destabilizing N-terminal residue of a substrate protein and participates in the formation of a substrate-linked multiubiquitin chain. We report the cloning of the mouse and human *Ubr1* cDNAs and genes that encode a mammalian N-recognin called E3α. Mouse UBR1p (E3α) is a 1,757-residue (200-kDa) protein that contains regions of sequence similarity to the 225-kDa Ubr1p of the yeast *Saccharomyces cerevisiae*. Mouse and human UBR1p have apparent homologs in other eukaryotes as well, thus defining a distinct family of proteins, the UBR family. The residues essential for substrate recognition by the yeast Ubr1p are conserved in the mouse UBR1p. The regions of similarity among the UBR family members include a putative zinc finger and RING-H2 finger, another zinc-binding domain. *Ubr1* is located in the middle of mouse chromosome 2 and in the syntenic 15q15-q21.1 region of human chromosome 15. Mouse *Ubr1* spans ≈120 kilobases of genomic DNA and contains ≈50 exons. *Ubr1* is ubiquitously expressed in adults, with skeletal muscle and heart being the sites of highest expression. In mouse embryos, the *Ubr1* expression is highest in the branchial arches and in the tail and limb buds. The cloning of *Ubr1* makes possible the construction of *Ubr1*-lacking mouse strains, a prerequisite for the functional understanding of the mammalian N-end rule pathway.

A number of regulatory circuits involve metabolically unstable proteins. Short *in vivo* half-lives are also characteristic of damaged or otherwise abnormal proteins (1–4). Features of proteins that confer metabolic instability are called degradation signals, or degrons. The essential component of one degradation signal, called the N-degron, is a destabilizing N-terminal residue of a protein (5, 6). The set of amino acid residues that are destabilizing in a given cell type yields a rule, called the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar, but distinct, versions of the N-end rule pathway are present in all organisms examined, from mammals to fungi and bacteria (6–8).

In eukaryotes, the N-degron comprises two determinants: a destabilizing N-terminal residue and an internal lysine or lysines (8). The Lys residue is the site of formation of a multiubiquitin chain (9). The N-end rule pathway is thus one pathway of the ubiquitin (Ub) system. Ub is a 76-residue protein whose covalent conjugation to other proteins plays a role in a multitude of processes, including cell growth, division, differentiation, and responses to stress (1, 3, 4, 10). In most of these processes, Ub acts through routes that involve the degradation of Ub-protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease (11).

The N-end rule is organized hierarchically. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their enzymatic deamidation into the secondary destabilizing N-terminal residues Asp and Glu (12). The destabilizing activity of N-terminal Asp and Glu requires their enzymatic conjugation to Arg, one of the primary destabilizing residues (6). The primary destabilizing N-terminal residues are bound directly by the *UBR1*-encoded N-recognin (also called E3), the recognition component of the N-end rule pathway (13). In *S. cerevisiae*, N-recognin is a 225-kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues—Phe, Leu, Trp, Tyr, Ile, Arg, Lys, and His. N-recognin has at least two substrate-binding sites. The type 1 site is specific for the basic N-terminal residues Arg, Lys, and His. The type 2 site is specific for the bulky hydrophobic N-terminal residues Phe, Leu, Trp, Tyr, and Ile (6).

The known functions of the N-end rule pathway include the control of peptide import in *S. cerevisiae* (through degradation of Cup9p, a transcriptional repressor of the peptide transporter Ptr2p); a role in controlling the Sln1p-dependent phosphorylation cascade that mediates osmoregulation in *S. cerevisiae*; the degradation of Gpa1p, a Gα protein of *S. cerevisiae*; and the degradation of alphaviral RNA polymerases in virus-infected metazoan cells (6, 14).

The mammalian counterpart of the yeast *UBR1*-encoded N-recognin (E3) was characterized biochemically in extracts from rabbit reticulocytes (15–17). Rabbit E3α was shown to be spe-

---

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/957898-6$2.00/0
PNAS is available online at http://www.pnas.org.

Abbreviations: Ub, ubiquitin; kb, kilobase; id., identity; si., similarity; BAC, bacterial artificial chromosome; FISH, fluorescence *in situ* hybridization; *e*n, embryonic day.

Data deposition: Nucleotide sequences reported in this work have been deposited in the GenBank database [accession nos. AF061555 (mouse *Ubr1* cDNA) and AF061556 (human *UBR1* cDNA)].

[**] To whom reprint requests should be addressed at: Division of Biology, 147-75, Caltech, 1200 East California Boulevard, Pasadena, CA 91125. e-mail: avarsh@cco.caltech.edu.

[††] The names of mouse genes are in italics, with the first letter uppercase. The names of human and *S. cerevisiae* genes are also in italics, all uppercase. If human and mouse genes are named in the same sentence, the mouse gene notation is used. The names of *S. cerevisiae* proteins are Roman, with the first letter uppercase and an extra lowercase "p" at the end. The names of the corresponding mouse and human proteins are the same, except that all letters but the last "p" are uppercase. The latter usage is a modification of the existing convention (33), to facilitate simultaneous discussions of yeast, mouse, and human proteins. In some citations, the abbreviated name of a species precedes the gene's name.

A

Peptide	Determined sequence	Position
T70B	YMSYSYDFR I S	324-332
T70A	FNFQGYSQDK	459-468
T98	GILISKPTISIER Y V WT	480-493
T120	QVGQHIEVDPDWEAAIAYQMQLK I	521-543
T100 (PEP3)	VSEDLVSINLPLSYSK H RTL	600-615
T134	FVPFEDFHIEVLVEYPLR DS VQ	634-651
T74B	VPQEFNVTKXEVTM GVG RE I	755-769
T76 (PEP2)	NLPENENNETGLENVINK	789-806
T92	FNMYFYHYSK	831-840
T96 (PEP1)	APEEEVVFDFYHK A	932-944
T122	GETLDPLFMDPDLAYGTY H	1140-1158
T74A	YFEAVQLSSQQR	1174-1185
T132	VDLFDLESGEYL	1188-1199
T140	QETNQMLFGFN	1743-1753

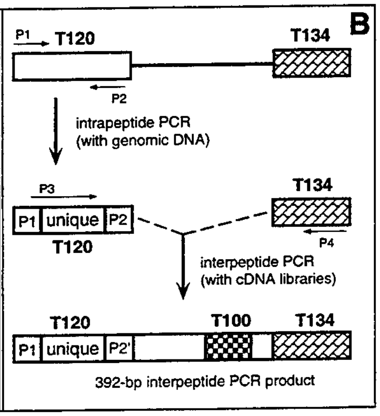


FIG. 1. Peptides of rabbit UBR1p (E3α) and isolation of the mouse *Ubr1* cDNA. (*A*) Amino acid sequences of tryptic peptides of the purified rabbit UBR1p (see *Materials and Methods*). The alternative sets of peptide names, T-based and PEP1-PEP3 (in parentheses), refer to two different preparations of E3α. The sequences of T120, T76, T96, and T122 that were encoded by DNA sequences identified through intrapeptide PCR are underlined. Residues deduced from the mouse *Ubr1* cDNA that differed from those inferred through peptide sequencing are indicated in a smaller font. The peptides' positions in the deduced sequence of mouse UBR1p are indicated. (*B*) The intrapeptide/interpeptide-PCR cloning strategy. The products of the initial intrapeptide PCR, derived from rabbit genomic DNA, were used to carry out interpeptide PCRwith a rabbit liver cDNA library (CLONTECH). The resulting 392-bp fragment of the rabbit *Ubr1* cDNA was used to isolate, using PCR and a λgt11 mouse liver cDNA library, the corresponding 392-bp mouse *Ubr1* cDNA fragment. This fragment then was used to screen the same cDNA library, yielding a 2.4-kb fragment of the mouse *Ubr1* cDNA that encoded several of the peptide-derived sequences of the rabbit UBR1p. The encoded sequence was also significantly similar to that of the N-terminal region of *S. cerevisiae* Ubr1p (13) and contained the putative start (ATG) codon of the mouse *Ubr1* ORF. To isolate the rest of the 5′ region of the *Ubr1* cDNA, 5′-rapid amplification of cDNA ends (RACE)–PCR (20) was performed with poly(A)$^+$ RNA from mouse L cells and a primer from the 2.4-kb DNA fragment. 3′-RACE–PCR (20) was used to amplify a downstream region of *Ubr1* cDNA. The resulting DNA fragment (nucleotides 2,470–3,467) then was used to screen a λgt10 mouse cDNA library from MEL-C19 cells. Five overlapping cDNA isolates (MR16, MR17, MR19, MR20, and MR23) that together spanned the entire *Ubr1* cDNA were mapped and subcloned into Bluescript II SK$^+$ (Stratagene), yielding the plasmid MR26, which contained the entire ORF of *Ubr1*. The ORF region of *Ubr1* cDNA was sequenced on both strands at least twice, using independently derived cDNA clones.

cifically required for the Ub-dependent degradation of proteins bearing either type 1 (basic) or type 2 (bulky hydrophobic) destabilizing N-terminal residues (7, 15, 16).

We began dissection of the mouse N-end rule pathway by isolating the *Ntan1* gene, which encodes the asparagine-specific N-terminal amidase (18, 19), a component of the mammalian N-end rule pathway, and by constructing mouse strains that lack *Ntan1* (Y.T.K. and A.V., unpublished data). Herein, we describe the cloning and characterization of the mouse and human cDNAs and genes[††] that encode UBR1p (E3α), a homolog of yeast Ubr1p and the main recognition component of the N-end rule pathway.

MATERIALS AND METHODS

Isolation and Partial Sequencing of Mammalian E3α (UBR1p). Rabbit E3α was purified from reticulocyte extracts by using affinity chromatography with immobilized protein substrates of UBR1p and elution with dipeptides bearing destabilizing N-terminal residues (16). The resulting preparation was fractionated by SDS/PAGE. The band of ≈180-kDa E3α was excised and subjected to digestion with trypsin. Amino acid sequences were determined for 14 peptides of rabbit UBR1p (Fig. 1*A*) by using standard methods (20).

Isolation of the Full-Length Mouse *Ubr1* cDNA. A strategy that included the intrapeptide-interpeptide PCR (21) was used (see the legend to Fig. 1).

Isolation of a Partial Human *UBR1* cDNA. Poly(A)$^+$ RNA from human 293 cells was subjected to reverse transcription–PCR, using sets of primers corresponding to sequences of the mouse *Ubr1* cDNA. One of the reactions yielded a 1.0-kilobase (kb) fragment that encompassed a region of the human *UBR1* cDNA (Fig. 2).

Mouse and Human Genomic *Ubr1* Fragments. A library of mouse genomic DNA fragments (strain SvJ) in bacterial artificial chromosome (BAC) (22) vector (Genome Systems, St. Louis) was used, as described in the legend to Fig. 2.

Northern, Southern, and Whole-Mount *in Situ* Hybridizations. Mouse and human multiple-tissue Northern blots

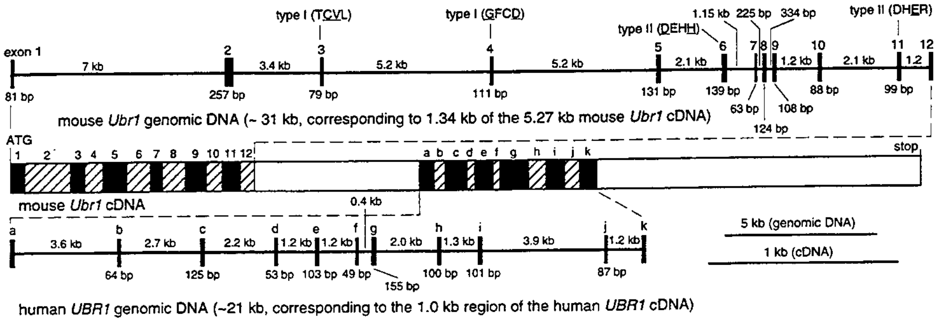


FIG. 2. The mouse and human *Ubr1* cDNAs and genes. Thick horizontal lines represent genomic DNA. The upper one is a ≈31-kb fragment of the mouse *Ubr1* gene that corresponds to a 1.34-kb region of mouse *Ubr1* cDNA (nucleotides 115–1,454). Vertical rectangles represent exons. Their lengths, and the lengths of the introns, are indicated, respectively, below and above the horizontal line. In a composite diagram of the *Ubr1* cDNA, the exons are depicted as alternatively shaded rectangles. For exon 1, only its translated region is indicated. Shown belowthe cDNA diagram is a ≈21-kb fragment of the human *UBR1* gene, corresponding to 1.0 kb of the indicated region of the human *UBR1* cDNA (nucleotides 2,218–3,227 of the mouse *Ubr1* cDNA sequence). The mouse and human *Ubr1* exons are denoted, respectively, by numbers and letters. Also indicated are the exon locations of some of the type 1 and type 2 substrate-binding sites of N-recognin (the essential amino acid residues are underlined) (A. Webster, M. Ghislain, and A.V., unpublished data; see the main text). Not shown are the 114-bp 5′-untranslated region (UTR) and the 1,010 bp 3′-UTR of the mouse *Ubr1* cDNA. To isolate mouse *Ubr1*, a library of mouse genomic DNA fragments in a BAC vector (see *Materials and Methods*) was screened with a fragment of the mouse *Ubr1* cDNA (nucleotides 105–1,333) as a probe, yielding seven BAC clones, of which BAC3 and BAC4 contained the entire *Ubr1* gene. The exon/intron organization of the first 31 kb (≈1/4) of the mouse *Ubr1* gene was determined by using exon-specific PCR primers to produce ≈40 genomic DNA fragments of the BAC3 insert that ranged in size from 1.3 to 18 kb. Regions encompassing the exon/intron junctions then were sequenced by using intron-specific primers. Fragments of the human genomic *UBR1* DNA were isolated by using primers derived from the 1.0-kb fragment of the human *UBR1* cDNA, the Expand High Fidelity PCR System (Roche Molecular Biochemicals, Indianapolis, IN), and genomic DNA from human 293 cells. The resulting four fragments were subcloned into pCR2.1 (Invitrogen), yielding the plasmids HR8, HR6–4, HR2–25, and HR7–2, whose partially overlapping inserts encompassed ≈21 kb of the human *UBR1* gene. Partial sequencing of the mouse and human genomic *Ubr1* fragments (≈20 kb of sequenced DNA) included all of the exon/intron junctions in these regions of *Ubr1*.

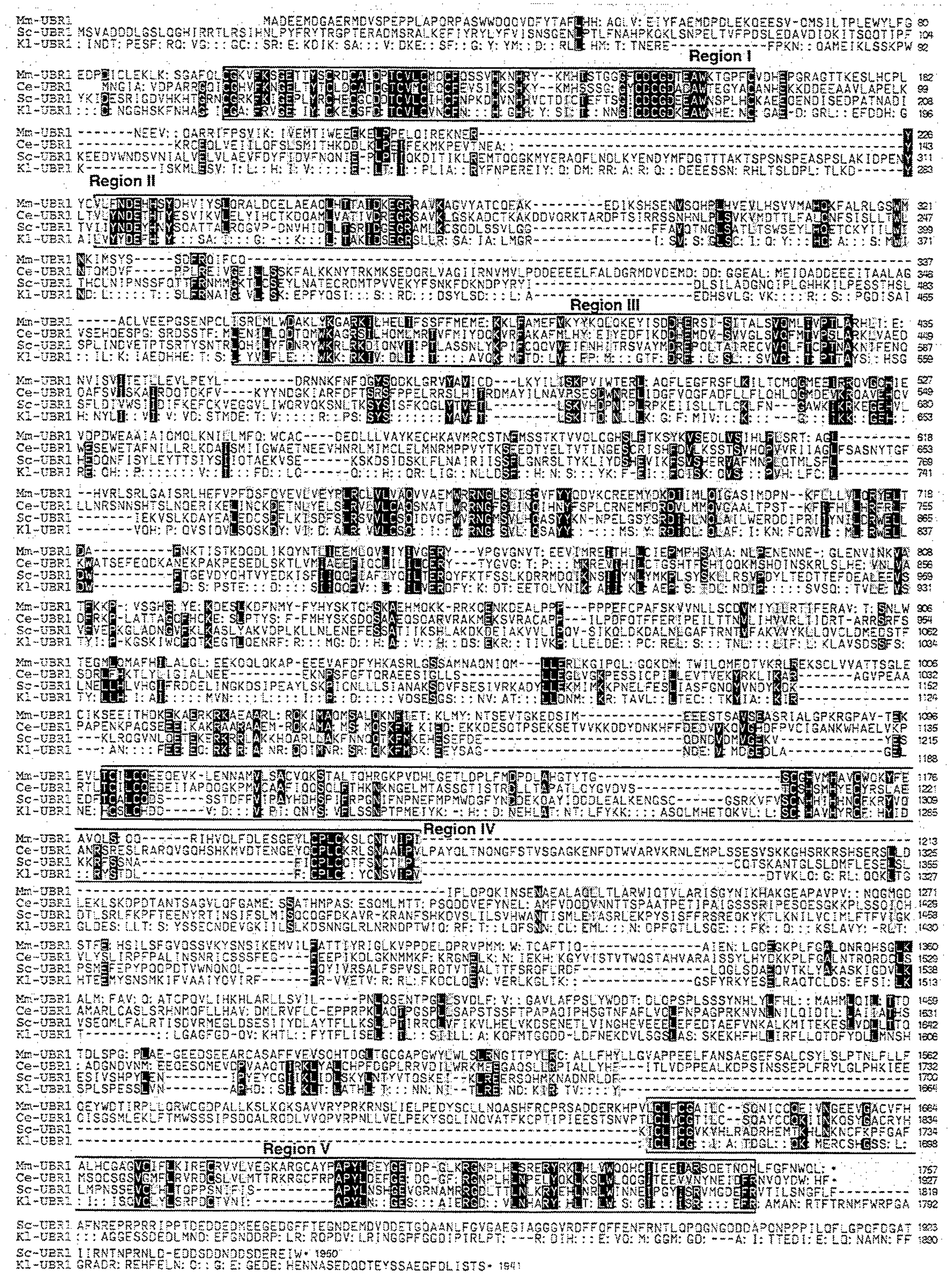


FIG. 3. Comparison of the deduced amino acid sequence of mouse UBR1p (Mm-UBR1) with those of *C. elegans* UBR1p (Ce-UBR1), *S. cerevisiae* Ubr1p (Sc-UBR1), and *K. lactis* Ubr1p (Kl-UBR1). White-on-black and gray shadings highlight, respectively, identical and similar residues. The residues of UBR proteins that are identical to those of *S. cerevisiae* Ubr1p are denoted by double dots, at positions where the identity involves just one non-*cerevisiae* protein. Also indicated are the regions of significant similarity among the four proteins. *K. lactis UBR1* was cloned through its crosshybridization to *S. cerevisiae UBR1* (P. Waller and A.V., unpublished data).

(CLONTECH), and either mouse or human *Ubr1* cDNA fragments labeled with ^{32}P were used (20). Southern hybridizations were carried out by using standard techniques (20). Mouse embryos were staged, fixed, and processed for *in situ* hybridization as described (23). For sectioning, the stained embryos were embedded in OCT medium (Sakura Finetek, Torrance, CA). A 1.2-kb *Ubr1* cDNA fragment (nucleotides 3,150–3,355) was used as a template for synthesizing antisense- or sense-strand RNA probes labeled with digoxigenin (23).

Chromosome Mapping of the Mouse and Human *Ubr1*. The mapping of mouse *Ubr1* was carried out by using the interspecific backcross analysis (24), essentially as described (18). Human *UBR1* was mapped by using fluorescence *in situ* hybridization

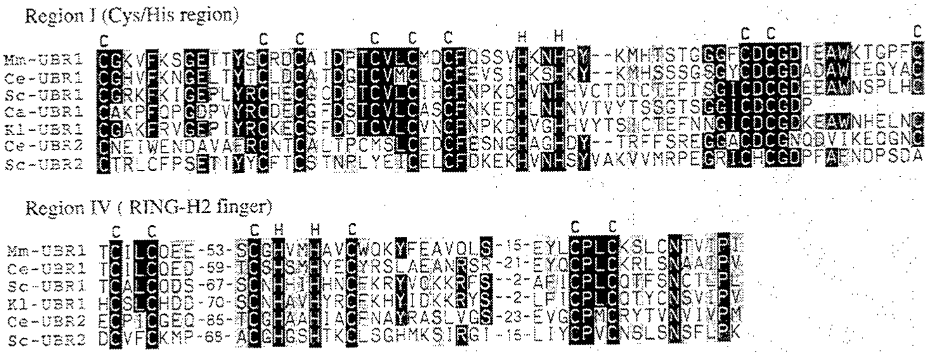


FIG. 4. Two Cys/His domains of the UBR protein family. Comparison of the putative zinc finger (region I) and RING-H2 finger (region IV) with the corresponding sequences from the other species in Fig. 3, and also with *C. albicans* Ubr1p (Ca-UBR1), *C. elegans* UBR2p (Ce-UBR2), and *S. cerevisiae* Ubr2p (Sc-UBR2). Numbers indicate the lengths of gaps. The conserved Cys and His residues are indicated.

(FISH) with mitotic chromosomes from human lymphocytes (25). The probe was a mixture of the HR8, HR6–4, HR2–25, and HR7–2 plasmids, labeled with biotin using biotinylated dATP and the BioNick labeling kit (Life Technologies, Grand Island, NY), and detected by using fluorescein isothiocyanate-avidin.

RESULTS AND DISCUSSION

Isolation of the Mouse *Ubr1* cDNA. Tryptic peptides of the purified rabbit E3α (UBR1p) (16) were isolated and sequenced (see *Materials and Methods*), yielding 14 short regions of E3α (Fig. 1*A*). These regions lacked significant similarities to the deduced sequence of *S. cerevisiae* Ubr1p (13). We used intrapeptide PCR (21) to identify a unique (nondegenerate) sequence of the rabbit *Ubr1* cDNA. This method allows amplification of a short unique DNA sequence by using two degenerate PCR primers (derived by reverse translation) that flank this sequence and correspond to the outermost regions of a single peptide (Fig. 1*B*). Several intrapeptide nucleotide sequences were obtained this way (Fig. 1). These sequences, together with those of the original degenerate primers, then were used to amplify a 392-bp fragment of the rabbit *Ubr1* cDNA, using interpeptide PCR (Fig. 1). This fragment encoded peptides T120 and T134 at either end and peptide T100 in the middle (Fig. 1*B*). A homologous 392-bp fragment of the mouse *Ubr1* cDNA then was amplified by using the same method (Fig. 1*B*). The rabbit and mouse 392-bp *Ubr1* cDNA fragments were 88% and 89% identical at the nucleotide and amino acid sequence levels, respectively, but lacked significant similarities to *S. cerevisiae UBR1* (data not shown).

The 392-bp mouse *Ubr1* cDNA fragment then was used, in conjunction with standard cDNA library screening and rapid amplification of cDNA ends–PCR (20), to isolate multiple *Ubr1* cDNA fragments, and to assemble them into a 5,271-bp ORF encoding a 1,757-residue protein (pI of 6.0), whose size, 200 kDa, was close to the estimated size of the isolated rabbit UBR1p (E3α), ≈180 kDa (16) (Figs. 2 and 3). The inferred ATG start codon (Fig. 2), within the sequence CTTAAGATGGCG, is preceded by two in-frame stop codons, at positions −48 and −93, and is located in a favorable Kozak context (26), with A and G at positions −3 and +4, respectively. There are two more ATGs, five and 11 codons downstream of the inferred one. These alternative start codons are in a favorable Kozak context as well.

Cloning and Partial Characterization of the Mouse and Human *Ubr1* Genes. A fragment of the mouse *Ubr1* cDNA was used to isolate a ≈120-kb mouse *Ubr1* genomic DNA clone, carried in a BAC vector (22). We determined the exon/intron organization and restriction map of the ≈31-kb region of *Ubr1* that corresponded to the 1,340-bp 5′-region of the mouse *Ubr1* cDNA (nucleotides 105–1,333) (Fig. 2). The lengths of the 12 exons in this region of mouse *Ubr1* range from 63 to 257 bp (Fig. 2).

The nucleotide and deduced amino acid sequences of the 1.0-kb human *UBR1* cDNA fragment (see *Materials and Methods*), located approximately in the middle of *UBR1* cDNA (nucleotides 2,218–3,227 of the mouse *Ubr1* cDNA) (Fig. 2), were, respectively, 91% and 94% identical to the corresponding mouse *Ubr1* cDNA and UBR1p sequences. Overlapping genomic DNA fragments of human *UBR1* that, together, encompassed a ≈21-kb region of the human *UBR1* gene and corresponded to the 1.0-kb fragment of the human *UBR1* cDNA (Fig. 2), were isolated from human DNA by using cDNA-derived primers and PCR. Partial sequencing showed that this ≈21-kb region of human *UBR1* contained 11 exons whose length ranged from 49 to 155 bp, a distribution of exon lengths similar to that in a different region of mouse *Ubr1* (Fig. 2). All of the sequenced exon/intron junctions (≈23 exons), which encompassed a ≈52-kb region of the mouse and human *Ubr1*, contained the consensus GT and AG dinucleotides characteristic of the mammalian nuclear pre-mRNA splice sites (data not shown) (20). Extrapolating from these data on the mouse and human *Ubr1* genes and the corresponding regions of their cDNAs (Fig. 2), a mammalian *Ubr1* gene is expected to be ≈120 kb long and to contain ≈50 exons.

The Mouse UBR1p Protein and its Homologs. The low overall sequence similarity of mouse UBR1p (E3α) to Ubr1p of either *S. cerevisiae* [22% identity (id.), 48% similarity (si.)] or another budding yeast, *Kluyveromyces lactis* (21% id., 48% si.), belied the presence of five regions, denoted I–V, which were significantly similar between the mouse and yeast versions of UBR1p (Figs. 3 and 4). By contrast, the *Ntan1*-encoded asparagine-specific N-terminal amidase, the most upstream component of the mouse N-end rule pathway, lacks sequence similarities to its *S. cerevisiae* counterpart Nta1p (12, 18). Database searches identified other likely homologs of mouse UBR1p, in particular the 1,927-residue protein of the nematode *Caenorhabditis elegans* (GenBank accession no. U88308) (32% id., 53% si.; termed *Ce-Ubr1*); the 1,872-residue *S. cerevisiae* protein (GenBank accession no. Z73196) (21% id., 47% si.; termed *Sc-UBR2*; ref. 4); the 2,168-residue *C. elegans* protein (GenBank accession no. U40029) (21% id., 45% si.; termed *Ce-Ubr2*); and the 794-residue CER3p protein of the plant *Arabidopsis thaliana* (GenBank accession no. X95962) (26% id., 49% si.). CER3p is involved in wax biosynthesis in *A. thaliana* (27). In addition, a 147-residue sequence of the yeast *Candida albicans* (http://alces.med.umn.edu/bin/genelist?LUBR1) was similar to the N-terminal region of mouse UBR1p (Fig. 4).

The presence of high-similarity regions I–V among these deduced sequences (Figs. 3 and 4) suggested the existence of a distinct protein family, termed UBR. The 66-residue region I, near the N terminus of UBR1p, is a particularly clear UBR family-identifying region (e.g., 61% id., 75% si. between mouse and *C. elegans* UBR1p) (Figs. 3 and 4).

Recent genetic analyses of *S. cerevisiae* Ubr1p (N-recognin) have shown that the regions I–III contain residues essential for the recognition of N-end rule substrates by Ubr1p. In particular, Cys-145, Val-146, Gly-173, and Asp-176 of region I were identified as essential residues of the type 1 binding site of *S. cerevisiae* Ubr1p (A. Webster, M. Ghislain, and A.V., unpublished data). All four of these residues were conserved between the yeast, mouse, and *C. elegans* UBR1p (Figs. 3 and 4). Region I is present in all of the known UBR family members except CER3p of *A. thaliana*, which contains only regions IV and V (Fig. 4). Region I encompasses a Cys/His-rich domain, Cys-X_{12}-Cys-X_2-Cys-X_5-

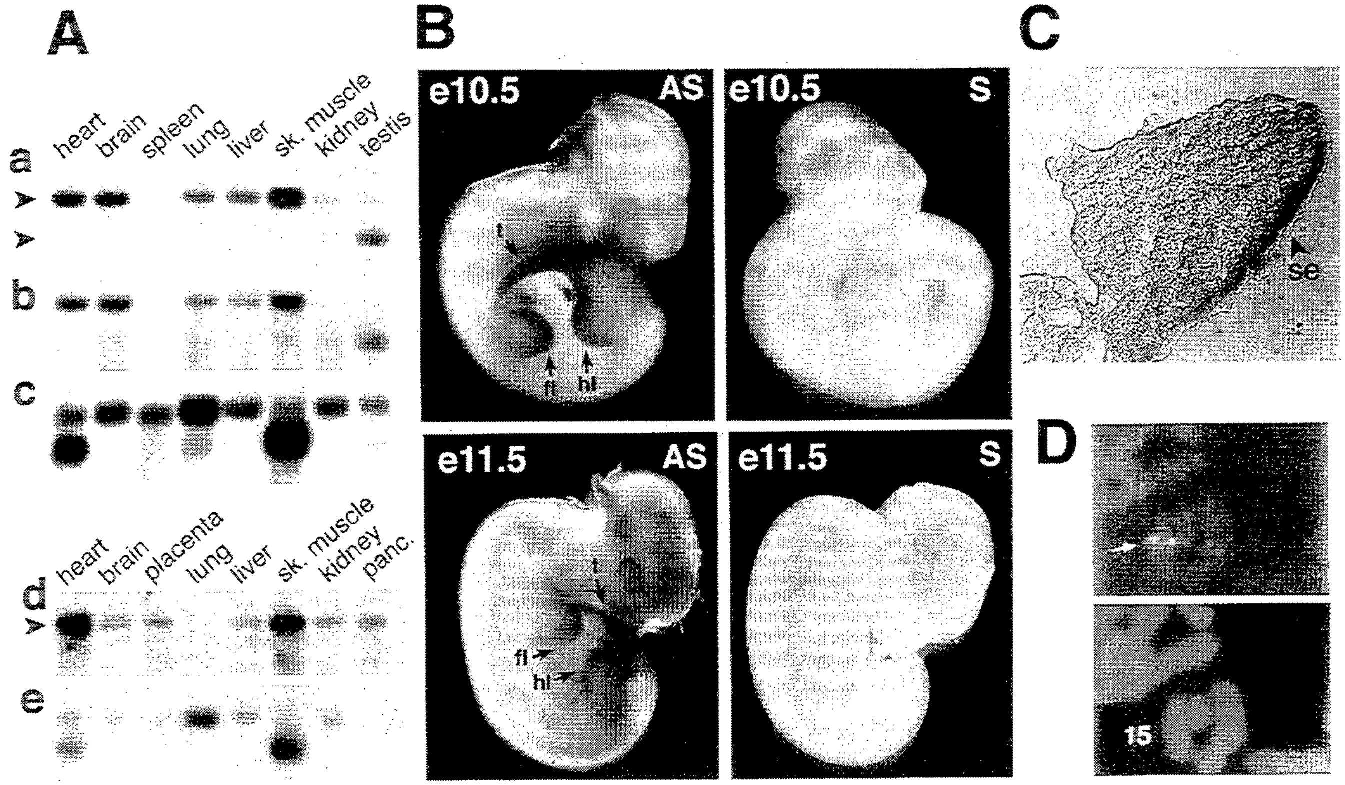


FIG. 5. Northern and *in situ* hybridizations with mouse and human *Ubr1*. (*A*) Membranes containing electrophoretically fractionated poly(A)$^+$ mRNA from different mouse (*a–c*) or human (*d* and *e*) tissues were hybridized with either a 2-kb 5′-proximal (nucleotides 116–2,124) mouse *Ubr1* cDNA fragment (*a*), its 0.64-kb 3′-proximal (nucleotides 4,749–5,388) fragment (*b*), a 1-kb human *UBR1* cDNA fragment (*d*), or the human β-actin cDNA fragment (*c* and *e*). The upper arrows in *a* and *d* indicate the ≈8-kb *Ubr1* transcript. The lower arrow in *a* indicates the ≈6-kb testis-specific *Ubr1* transcript. In the RNA sample from mouse spleen, the *Ubr1* transcript (but not the actin transcript) may have been degraded (*a–c*). (*B*) Expression of *Ubr1* in e10.5 and e11.5 mouse embryos. Whole-mount *in situ* hybridization was carried out with either antisense (AS) or sense (S, negative control) *Ubr1* cDNA probes (see *Materials and Methods*). The regions of high *Ubr1* expression are indicated by arrows (t, tail; fl, forelimb buds; hl, hindlimb buds). The branchial arches, where *Ubr1* is also highly expressed in e10.5 embryos (data not shown), are not visible in this e10.5 embryo. (*C*) Expression of *Ubr1* in the surface ectoderm of limb buds. Shown is a transverse section of a forelimb bud of an e10.5 embryo (se, surface ectoderm). (*D*) FISH analysis of human *UBR1*. (*Upper*) An example of the *UBR1*-specific FISH signal (arrow). (*Lower*) The same mitotic spread stained with 4′-6-diamino-2-phenylindole (DAPI) to visualize the chromosomes (see also Fig. 6).

Cys-X$_2$-Cys-X$_2$-Cys-X$_5$-His-X$_2$-His-X$_{(12–14)}$-Cys-X$_1$-Cys-X$_{11}$-Cys (Figs. 3 and 4), which is distinct from the known consensus sequences of zinc fingers and other Cys/His-motifs. Residues Asp-318, His-321, and Glu-560 of *S. cerevisiae* Ubr1p, which have been identified as essential for the type 2 binding site of this N-recognin (A. Webster, M. Ghislain, and A.V., unpublished data), were found to be retained in region II (Asp-318 and His-321) and region III (Glu-560) of the mouse and *C. elegans* UBR1p (Fig. 3).

Region IV contains another Cys/His-rich domain of UBR1p, Cys-X$_2$-Cys-*loop 1*-Cys-X$_1$-His-X$_2$-His-X$_2$-Cys-*loop 2*-Cys-X$_2$-Cys (Figs. 3 and 4), which is present in all of the UBR family members, and fits the consensus sequence of the RING-H2 finger, a subfamily of the previously defined RING motif (28). At least some of the RING-H2 sequences are sites of specific protein–protein interactions (28). Apc11p, a subunit of the Ub-protein ligase complex called the cyclosome (2) or the anaphase promoting complex, also contains a RING-H2 finger (29).

Another area of similarity (24–50% id., 46–70% si.) among the UBR family members is region V (Fig. 3 and data not shown). This region, 115 residues long in mouse UBR1p, near the protein's C terminus, is particularly similar between mouse and *C. elegans* UBR1p (50% id., 70% si.) (Fig. 3). Region V is located 4–14 residues from the UBR proteins' C termini, the exceptions being the *S. cerevisiae* and *K. lactis* Ubr1p, which bear, respectively, 132- and 159-residue tails of unknown function that are rich in the acidic Asp/Glu residues (36% and 33%) (Fig. 3). No significant similarities could be detected between mammalian UBR1p and other E3s (recognins) of the metazoan Ub system, including E6AP (30) and subunits of the cyclosome/anaphase promoting complex, except for the presence of a RING-H2 finger domain in the latter (29). [Different E3 proteins of the Ub system recognize different degrons in protein substrates, thereby defining distinct Ub-dependent proteolytic pathways (1, 4).]

Expression of Mouse and Human *Ubr1*. The 5′- and 3′-proximal mouse cDNA probes yielded similar results, detecting a single ≈8-kb transcript in several tissues (Fig. 5 *Aa* and *Ab*). In the testis, however, the ≈8-kb species of *Ubr1* mRNA was a minor one, the major species being ≈6 kb (Fig. 5*Aa*). The levels of either mouse or human *Ubr1* mRNA were highest in skeletal muscle and heart (Fig. 5*A*). The expression of mRNA encoding E2$_{14K}$, one of the mouse Ub-conjugating (E2) enzymes and a likely component of the mouse N-end rule pathway (6), was also highest in skeletal muscle and heart (18).

The distinct *Ubr1* mRNA pattern in the testis (Fig. 5*A*) was reminiscent of the analogous expression pattern of *Ntan1* mRNA, which encodes the Asn-specific N-terminal amidase, another component of the mammalian N-end rule pathway. Specifically, the size of the major species of *Ntan1* mRNA was ≈1.4 kb in all of the examined mouse tissues except testis, where the major species was ≈1.1 kb (18). The ≈1.1-kb *Ntan1* transcript recently was found to hybridize only to the 3′-half (exons 6–10 but not exons 1–5) of the *Ntan1* ORF (Y.T.K. and A.V., unpublished data). The functional significance of the testis-specific *Ubr1* and *Ntan1* expression patterns remains to be understood.

We used whole-mount *in situ* hybridization to examine the expression of *Ubr1* during embryogenesis. In e9.5 (9.5 days old)

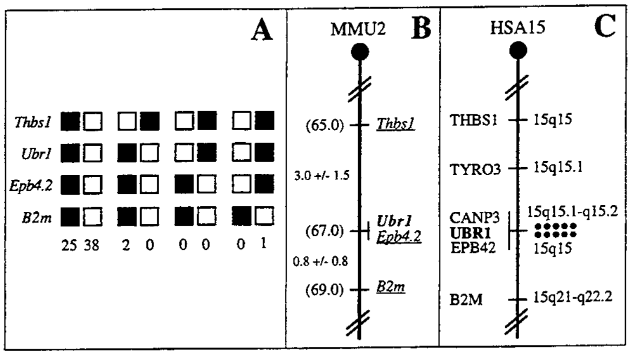


FIG. 6. Chromosomal locations of the mouse and human *Ubr1* genes. (*A*) Mouse *Ubr1* was mapped to the middle of mouse chromosome 2 by using interspecific (*M. musculus-M. spretus*) backcross analysis (18, 24). Shown are the segregation patterns of mouse *Ubr1* and the flanking genes in 66 backcross animals that were typed for all loci. For individual pairs of loci, more than 66 animals were typed. Each column represents the chromosome identified in the backcross progeny that was inherited from the [*M. musculus* C57BL/6J × *M. spretus*] F_1 parent. Filled and empty squares represent, respectively, C57BL/6J and *M. spretus* alleles. The numbers of offspring that inherited each type of chromosome 2 are listed below the columns. (*B*) A partial mouse chromosome 2 linkage map (MMU2), showing *Ubr1* in relation to the linked genes *Thbs1*, *Ebp4.2*, and *B2m*, and also, on the left, the corresponding recombination distances between the loci, in centimorgans, and the map locations, in parentheses. (*C*) A partial human chromosome 15 linkage map (HSA15). Each dot on the right, in the 15q15-q21.1 region, corresponds to the actually observed *UBR1*-specific double-dot FISH signal detected on human chromosome 15 (see also Fig. 5*D*).

mouse embryos, the expression of *Ubr1* was highest in the branchial arches and in the buds of forelimbs and the tail (data not shown). In e10.5 embryos, the expression of *Ubr1* became high in the hindlimb buds as well (Fig. 5*B*). This pattern was maintained in the limb buds of e11.5 embryos (Fig. 5*B*). High expression of *Ubr1* in the limb buds was confined predominantly to the surface ectoderm (Fig. 5*C*). This pattern of *Ubr1* expression in embryos (Fig. 5 *B* and *C*) is similar, if not identical, to that of *Ntan1*, which encodes asparagine-specific N-terminal amidase (18) (Y.T.K. and A.V., unpublished data), consistent with UBR1p and NTAN1p being components of the same pathway.

The enhanced expression of *Ubr1* in the embryonic limb buds (Fig. 5 *B* and *C*) is interesting in view of the conjecture that the N-end rule pathway might be required for limb regeneration in amphibians (31). The injection of dipeptides bearing destabilizing N-terminal residues into the stumps of amputated forelimbs of the newt was observed to delay limb regeneration, whereas the injection of dipeptides bearing stabilizing N-terminal residues had no effect (31). Rigorous tests of this and other suggested functions of the metazoan N-end rule pathway (6) will require mouse strains that lack *Ubr1*.

Chromosome Mapping of Mouse and Human *Ubr1*. The chromosomal location of mouse *Ubr1* was determined by interspecific backcross analysis, using DNA derived from matings of [(C57BL/6J × *Mus spretus*)F_1 × C57BL/6J] mice (Fig. 6 *A* and *B*) (18, 24). Mouse *Ubr1* is located in the central region of chromosome 2 and is linked to the *Thbs1*, *Epb4.2*, and *B2m* genes, the most likely gene order being centromere-*Thbs1*-*Ubr1*-*Epb4.2*-*B2m* (Fig. 6*B* and data not shown).

The chromosomal location of human *UBR1* was determined by using FISH (25), with human *UBR1* genomic DNA fragments as probes (Figs. 5*D* and 6*C*). This mapping placed *UBR1* at the 15q15–15q21.1 region of the human chromosome 15, an area syntenic with the independently mapped position of mouse *Ubr1* (Fig. 6). *Ubr1* is located in the regions of human chromosome 15 and mouse chromosome 2 that appear to be devoid of the previously mapped but uncloned mutations. Mutations in the human gene *CANP3*, which encodes a subunit of calpain and is located very close, if not adjacent, to *UBR1*, have been shown to cause a myopathy called the limb-girdle muscular distrophy (32).

Concluding Remarks. Isolation of the mouse and human *Ubr1* cDNAs and genes (Figs. 2–6) should enable functional understanding of the mammalian N-end rule pathway, in part through the construction and analysis of mouse strains that lack *Ubr1*. Recent searches in GenBank identified several mouse and human sequences in expressed sequence tag databases that exhibited significant similarity to the C-terminal region of mouse UBR1p. The cloning and characterization of the corresponding cDNAs have shown that there exist at least two distinct mouse (and human) genes, termed *Ubr2* and *Ubr3*, which encode proteins that are significantly similar to mouse UBR1p (Y.T.K. and A.V., unpublished data). Molecular and functional analyses of these *Ubr1* homologs are under way.

We are grateful to A. Webster and M. Ghislain for permission to cite their unpublished data. We thank members of the Varshavsky lab, especially I. V. Davydov, for helpful discussions, and L. Peck, G. Turner, H. Rao, A. Kashina, and F. Du for comments on the manuscript. Y.T.K. thanks B. Yu for sharing his Northern hybridization data on human β-actin mRNA. We gratefully acknowledge the sequencing of *K. lactis UBR1* by P. Waller. N.G.C. and N.A.J. thank D. J. Gilbert and D. B. Householder for excellent technical assistance. D.K.G. was a Scholar of the Leukemia Society of America. This study was supported by National Institutes of Health grants to A.V. (DK39520 and GM31530), V.A.F. (NS29542), and D.K.G. (GM45314), and by a grant to N.G.C. from the National Cancer Institute.

1. Varshavsky, A. (1997) *Trends Biochem. Sci.* **22,** 383–387.
2. Hershko, A. (1997) *Curr. Opin. Cell. Biol.* **9,** 788–799.
3. Haas, A. J. & Siepman, T. J. (1997) *FASEB J.* **11,** 1257–1268.
4. Hochstrasser, M. (1996) *Annu. Rev. Genet.* **30,** 405–439.
5. Bachmair, A., Finley, D. & Varshavsky, A. (1986) *Science* **234,** 179–186.
6. Varshavsky, A. (1997) *Genes Cells* **2,** 13–28.
7. Gonda, D. K., Bachmair, A., Wünning, I., Tobias, J. W., Lane, W. S. & Varshavsky, A. (1989) *J. Biol. Chem.* **264,** 16700–16712.
8. Bachmair, A. & Varshavsky, A. (1989) *Cell* **56,** 1019–1032.
9. Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K. & Varshavsky, A. (1989) *Science* **243,** 1576–1583.
10. Pickart, C. M. (1997) *FASEB J.* **11,** 1055–1066.
11. Baumeister, W., Walz, J., Zühl, F. & Seemüller, E. (1998) *Cell* **92,** 367–380.
12. Baker, R. T. & Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 12065–12074.
13. Bartel, B., Wünning, I. & Varshavsky, A. (1990) *EMBO J.* **9,** 3179–3189.
14. Byrd, C., Turner, G. C. & Varshavsky, A. (1998) *EMBO J.* **17,** 269–277.
15. Reiss, Y., Kaim, D. & Hershko, A. (1988) *J. Biol. Chem.* **263,** 2693–269.
16. Reiss, Y. & Hershko, A. (1990) *J. Biol. Chem.* **265,** 3685–3690.
17. Hershko, A. & Ciechanover, A. (1992) *Annu. Rev. Biochem.* **61,** 761–807.
18. Grigoryev, S., Stewart, A. E., Kwon, Y. T., Arfin, S. M., Bradshaw, R. A., Jenkins, N. A., Copeland, N. J. & Varshavsky, A. (1996) *J. Biol. Chem.* **271,** 28521–28532.
19. Stewart, A. E., Arfin, S. M. & Bradshaw, R. A. (1995) *J. Biol. Chem.* **270,** 25–28.
20. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G. & Struhl, K. (1996) *Current Protocols in Molecular Biology* (Wiley Interscience, New York).
21. Bredt, D. S., Hwang, P. M., Glatt, C. E., Lowenstein, C., Reed, R. R. & Snyder, S. H. (1991) *Nature (London)* **351,** 714–718.
22. Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y. & Simon, M. I. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8794–8797.
23. Conlon, R. A. & Rossant, J. (1992) *Development (Cambridge, U.K.)* **116,** 357–368.
24. Copeland, N. G. & Jenkins, N. A. (1991) *Trends Genet.* **7,** 113–118.
25. Dracopoli, N. C., Haines, J. L., Korf, B. R., Moir, T. D., Morton, C. C., Seidman, C. E., Seidman, J. G. & Smith, D. R. (1994) *Current Protocols in Human Genetics* (Wiley Interscience, New York).
26. Kozak, M. (1996) *Mamm. Genome* **7,** 563–574.
27. Hannoufa, A., Negruk, V., Eisner, G. & Lemieux, B. (1996) *Plant J.* **10,** 459–467.
28. Borden, K. L. & Freemont, P. S. (1996) *Curr. Opin. Struct. Biol.* **6,** 395–401.
29. Yu, H., Peters, J. M., King, R. W., Page, A. M., Hieter, P. & Kirschner, M. W. (1998) *Science* **279,** 1219–1222.
30. Huibregtse, J. M., Scheffner, M., Beaudenon, S. & Howley, P. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 2563–2567.
31. Taban, C. H., Hondermarck, H., Bradshaw, R. A. & Boilly, B. (1996) *Experientia* **52,** 865–870.
32. Richard, I., Broux, O., Allamand, V., Fougerousse, F., Chiannilkulchai, N., Bourg, N., Brenguier, L., Devaud, C., Pasturaud, P., Roudaut, C., *et al.* (1995) *Cell* **81,** 27–40.
33. Stewart, A. (1995) *Trends in Genetics Nomenclature Guide* (Elsevier, Cambridge, U.K.).

The EMBO Journal Vol.17 No.1 pp.269–277, 1998

The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor

Christopher Byrd, Glenn C.Turner and Alexander Varshavsky[1]

Division of Biology, California Institute of Technology,
1200 East California Boulevard, Pasadena, CA 91125, USA

[1]Corresponding author
e-mail: avarsh@cco.caltech.edu

Ubiquitin-dependent proteolytic systems underlie many processes, including the cell cycle, cell differentiation and responses to stress. One such system is the N-end rule pathway, which targets proteins bearing destabilizing N-terminal residues. Here we report that Ubr1p, the main recognition component of this pathway, regulates peptide import in the yeast *Saccharomyces cerevisiae* through degradation of Cup9p, a 35 kDa homeodomain protein. Cup9p was identified using a screen for mutants that bypass the previously observed requirement for Ubr1p in peptide import. We show that Cup9p is a short-lived protein ($t_{1/2}$ ~5 min) whose degradation requires Ubr1p. Cup9p acts as a repressor of *PTR2*, a gene encoding the transmembrane peptide transporter. In contrast to engineered N-end rule substrates, which are recognized by Ubr1p through their destabilizing N-terminal residues, Cup9p is targeted by Ubr1p through an internal degradation signal. The Ubr1p–Cup9p–Ptr2p circuit is the first example of a physiological process controlled by the N-end rule pathway. An earlier study identified Cup9p as a protein required for an aspect of resistance to copper toxicity in *S.cerevisiae*. Thus, one physiological substrate of the N-end rule pathway functions as both a repressor of peptide import and a regulator of copper homeostasis.

Keywords: *CUP9*/N-end rule/peptide import/proteolysis/*PTR2*/*UBR1*

Introduction

Many regulatory proteins are short-lived *in vivo* (Schwob *et al.*, 1994; King *et al.*, 1996; Varshavsky, 1996). This metabolic instability makes possible rapid adjustment of the protein's concentration (or subunit composition) through changes in the rates of its synthesis or degradation. Protein degradation plays a role in a multitude of processes, including cell growth, division, differentiation and responses to stress. In eukaryotes, a large fraction of intracellular proteolysis is mediated by the ubiquitin system. Ubiquitin (Ub) is a 76-residue protein whose covalent conjugation to other proteins marks them for processive degradation by the 26S proteasome—an ATP-dependent, multisubunit protease (Jentsch and Schlenker, 1995; Hilt and Wolf, 1996; Hochstrasser, 1996; Rubin *et al.*, 1997).

Features of proteins that confer metabolic instability are called degradation signals (degrons). One of the degradation signals recognized by the ubiquitin system is called the N-degron. It comprises two essential determinants: a destabilizing N-terminal residue and an internal lysine of a substrate (Bachmair *et al.*, 1986; Varshavsky, 1996). The Lys residue is the site of formation of a substrate-linked multiubiquitin chain (Bachmair and Varshavsky, 1989; Chau *et al.*, 1989). A set of N-degrons bearing different N-terminal residues that are destabilizing in a given cell type yields a rule, called the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar but distinct versions of the N-end rule operate in all organisms examined, from mammals to fungi and bacteria (Varshavsky, 1996).

The N-end rule pathway is organized hierarchically. In eukaryotes such as *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by the *NTA1*-encoded N-terminal amidase (Nt-amidase), into the secondary destabilizing residues Asp and Glu (Baker and Varshavsky, 1995). Secondary residues, in turn, function through their conjugation to Arg by the *ATE1*-encoded Arg-tRNA–protein transferase (R-transferase) (Balzi *et al.*, 1990). Arg is one of several primary destabilizing N-terminal residues which are bound directly by N-recognin, a 225 kDa E3 protein encoded by the *UBR1* gene (Bartel *et al.*, 1990). Ubr1p, together with the associated ubiquitin-conjugating (E2) enzyme Ubc2p, mediates the formation of a substrate-linked multiubiquitin chain (Varshavsky, 1996).

The N-end rule pathway was first encountered in experiments that explored, in *S.cerevisiae*, the metabolic fate of a fusion between Ub and a reporter such as *Escherichia coli* β-galactosidase (β-gal) (Bachmair *et al.*, 1986). While such engineered N-end rule substrates have been extensively characterized (Varshavsky, 1996), little is known about their physiological counterparts. The few identified so far include the *GPA1*-encoded Gα subunit of the *S.cerevisiae* heterotrimeric G protein, which mediates the pheromone response in this fungus, and RNA polymerases of alphaviruses whose hosts include mammalian and insect cells (de Groot *et al.*, 1991; Madura and Varshavsky, 1994). Physiological functions of the instability of these proteins remain to be understood (Varshavsky, 1996). Inactivation of the N-end rule pathway in *S.cerevisiae*—through deletion of the *UBR1* gene—results in cells which grow slightly slower than their wild-type counterparts, and are impaired in sporulation (increased frequency of asci containing fewer than four spores), but otherwise appear to be normal (Bartel *et al.*, 1990).

Recently, Becker and colleagues (Alagramam *et al.*, 1995) have reported that *ubr1*Δ cells are deficient in the

© Oxford University Press

A

Strain	Genotype	Peptide import	Strain	Genotype	Peptide import
JD53	wt	+	MHY599	*pas2Δ*	+
JD83-1A	*ubr1Δ*	–	CBY8	*ubc2Δ, ubc1Δ*	+
KM207-1	*ate1Δ*	+	CBY11	*ubc2Δ, cdc34-1*^{1}	+
SGY6	*nta1Δ*	+	EJY105/6	*ubc2Δ, ubc4Δ*	–
CBY20	*ptr2Δ*	+	CBY4	*ubc2Δ, ubc5Δ*	+
EJ2	*ubc1Δ*	+	CBY5	*ubc2Δ, ubc6Δ*	+
BBY67	*ubc2Δ*	+/–	CBY6	*ubc2Δ, ubc7Δ*	+
JD34	*ubc2Δ*	+/–	CBY7	*ubc2Δ, ubc8Δ*	+
EJY102	*ubc2Δ*	+/–	CBY9	*ubc2Δ, ubc9-1*^{1}	+
RJD549	*cdc34-1*^{1}	+	CBY10	*ubc2Δ, pas2Δ*	+
RJD795	*cdc34-2*^{1}	+	EJ10	*ubc4Δ, ubc5Δ*	+
EJ3	*ubc4Δ*	+	MHY503	*ubc4Δ, ubc6Δ*	+
EJ1	*ubc5Δ*	+	MHY554	*ubc4Δ, ubc7Δ*	+
MHY495	*ubc6Δ*	+	MHY552	*ubc6Δ, ubc7Δ*	+
MHY507	*ubc7Δ*	+	MHY550	*ubc4Δ, 5Δ, 6Δ*	+
MHY601	*ubc8Δ*	+	MHY557	*ubc4Δ, 6Δ, 7Δ*	+
YWO55	*ubc9-1*^{1}	+	MHY570	*ubc4Δ, 5Δ, 6Δ, 7Δ*	+

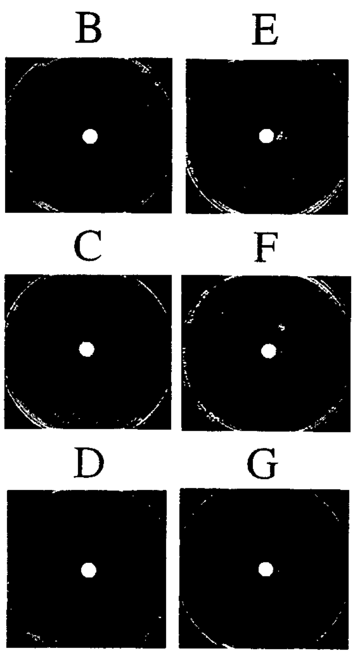


Fig. 1. The import of peptides is decreased in the absence of Ubc2p and virtually abolished in the absence of Ubc2p and Ubc4p. *Saccharomyces cerevisiae* mutants deficient in one or more ubiquitin-conjugating (E2) enzymes or other components of the ubiquitin system were tested for their ability to import peptides, using the halo assay and a toxic dipeptide L-leucyl-L-ethionine (Leu-Eth) (see Materials and methods). '+', '±' and '-' denote, respectively, the apparently wild-type, significantly reduced, and undetectable levels of peptide import. A superscript '1' refers to strains that carried a *ts* allele of an essential E2 enzyme, either Cdc34p (Ubc3p) or Ubc9p; the 30°C temperature of the test was semi-permissive for these strains. (**A**) Summary of the results. (B–F) Examples of the actual halo assays, with wild-type (**B**), *ptr2Δ* (**C**), *ubr1Δ* (**D**), *ubc2Δ* (**E**), *ubc4Δ* (**F**) and *ubc2Δ ubc4Δ* (**G**) strains of *S.cerevisiae*. Elimination of some E2 enzymes in the *ubc2Δ* background restored halo formation, presumably because such strains were growth-impaired in a way that made them hypersensitive to the toxicity of ethionine. By contrast, although *ubc2Δ ubc4Δ* cells were also growth-impaired, they were import-defective and therefore grew in the immediate vicinity of the filter.

import of di- and tripeptides, suggesting that this process, which is universal among living cells, requires the N-end rule pathway. In the present work, we identified the underlying regulatory mechanism and discovered a new physiological substrate of the N-end rule pathway, the homeodomain protein Cup9p. This short-lived protein is targeted for degradation by Ubr1p, and acts as a transcriptional repressor of *PTR2*, a gene that encodes a transmembrane peptide transporter. The Ubr1p–Cup9p–Ptr2p circuit is the first example of a physiological process controlled by the N-end rule pathway.

Results

The involvement of Ubc2p and Ubc4p E2 enzymes in the control of peptide import

We began by asking whether components of the N-end rule pathway other than Ubr1p were also necessary for the import of peptides. Previous work has shown that Ubc2p, one of 13 ubiquitin-conjugating (E2) enzymes of *S.cerevisiae*, is required for the degradation of engineered N-end rule substrates, and is physically associated with the E3 protein Ubr1p (N-recognin) (Jentsch, 1992; Madura *et al.*, 1993).

To test for the ability of *S.cerevisiae* to import peptides, we used a halo assay, in which a filter soaked in the toxic dipeptide L-leucyl-L-ethionine (Leu-Eth) is placed on a plate, inhibiting the growth of import-competent cells near the filter. By this test, the elimination of *UBC2* impaired, but did not abolish, the import of peptides (Figure 1A, B and E). To determine which of the other E2 enzymes, if any, were required for the residual peptide import observed in *ubc2Δ* cells, a number of single and multiple mutants in *UBC* genes were examined (Figure 1). We found that the elimination of both *UBC2* and *UBC4* virtually abolished the import of peptides (Figure 1A, B and E–G). Elimination of Ubc4p, one of the more abundant E2 enzymes (Bachmair *et al.*, 1986; Jentsch, 1992), had previously been noticed to decrease slightly the activity of the N-end rule pathway (Bartel, 1990). Cells lacking either Nta1p or Ate1p—the 'upstream' components of this pathway—were also examined and found unimpaired in the import of peptides, in contrast to *ubr1Δ* and *ubc2Δ ubc4Δ* cells (Figure 1A).

Identification of Cup9p as a negative regulator of peptide import

Previous work (Alagramam *et al.*, 1995) has shown that deletion of *UBR1* greatly reduces the level of *PTR2* mRNA, which encodes the transmembrane peptide transporter. This result, and our observation that peptide import requires the presence of at least one of two specific ubiquitin-conjugating enzymes (Figure 1), suggested a model in which expression of the Ptr2p transporter is regulated by a short-lived repressor that is degraded by the N-end rule pathway. In cells lacking *UBR1*, the repressor would be expected to accumulate, thereby blocking peptide import. One prediction of this model is that inactivation of this repressor would bypass the requirement for Ubr1p in peptide import. A screen for such 'bypass' mutations (see Materials and methods) yielded 199 recessive isolates, of which 101 defined one complementation group, termed

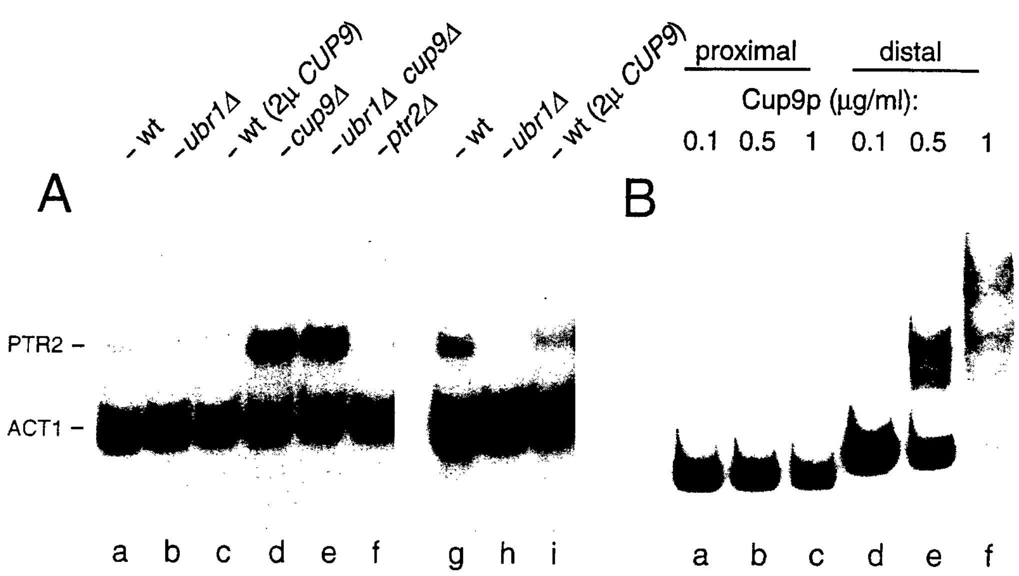


Fig. 2. Cup9p is a repressor of the *PTR2* gene. (**A**) Expression of the *S.cerevisiae PTR2* gene in different genetic backgrounds. Equal amounts of total RNA isolated from different strains were analyzed by Northern hybridization (see Materials and methods), using the *PTR2* (peptide transporter) and *ACT1* (actin) genes as ^{32}P-labeled probes. Lane a, JD52 (*CUP9 UBR1*) (wild-type) transformed with pCB201 (empty vector). Lane b, JD55 (*CUP9 ubr1Δ*) transformed with pCB201. Lane c, JD52 (*CUP9 UBR1*) transformed with the high-copy plasmid pCB209 that expressed *CUP9* from its natural promoter. Lane d, CBY18 (*cup9Δ UBR1*). Lane e, CBY16 (*cup9Δ ubr1Δ*). Lane f, CBY21 (*CUP9 UBR1 ptr2Δ*). Lanes g–i, same as lanes a–c, but a longer autoradiographic exposure to highlight the tight repression of *PTR2* by Cup9p in *ubr1Δ* cells (b, h) and the difference between levels of *PTR2* mRNA in the wild-type cells (a, g) and their counterparts that overexpressed Cup9p (c, i). (**B**) Cup9p specifically binds to a site in the P_{PTR2} promoter. A gel shift assay with Cup9-H$_6$, poly-dI·dC and ^{32}P-labeled DNA fragments of the *PTR2* promoter hybridization (see Materials and methods). Lanes a–c, with a fragment (–1 to –447) proximal to the inferred start codon of the *PTR2* ORF. Lanes d–f, same as lanes a–c, but with a more distal DNA fragment (–448 to –897). Concentrations of Cup9-H$_6$ (in µg/ml) are indicated above the lanes.

sub1 (suppressor of a block to peptide import in *ubr1Δ*). In agreement with the model's prediction, *sub1* mutants acquired the ability to express *PTR2* in the absence of *UBR1* (Figure 2A, lane b versus lane e). The *sub1* locus was cloned by complementation (see Materials and methods), and was found to be the *CUP9* gene.

CUP9 was originally identified by Knight *et al.* (1994) as a gene whose disruption impairs the copper resistance of *S.cerevisiae* growing on lactate, a non-fermentable carbon source. Under these conditions, Cup9p plays a major (but mechanistically obscure) role in copper homeostasis (Knight *et al.*, 1994). *CUP9* encodes a 35 kDa protein that contains a homeodomain, an ~60-residue helix–turn–helix DNA-binding motif present in many eukaryotic regulatory proteins (Wolberger, 1996). Outside the homeodomain region, the sequence of Cup9p is not similar to sequences in databases.

To verify that *CUP9* and *SUB1* were the same gene, complementation tests were carried out. Two independently derived *ubr1Δ cup9::LEU2* strains (CBY16 and CBY17) were crossed to *ubr1Δ sub1-1* (CBY15), and the resulting diploids (CBY23 and CBY24, respectively) were tested for their ability to import dipeptides (see Materials and methods). As would be expected of allelic loci, *cup9::LEU2* and *sub1-1* failed to complement one another: both diploid strains remained import-competent (data not shown). In another test, CBY23 and CBY24 were sporulated, and the segregants were analyzed for the presence of the *LEU2* gene (integrated at the *CUP9* locus) and for the ability to import peptides. Among the eight tetrads tested, *LEU2* was present in two of the four segregants, whereas all four segregants were invariably import-competent, a pattern expected if *CUP9* and *SUB1* were one and the same gene.

To examine the regulation of peptide import by Cup9p and Ubr1p, congenic *S.cerevisiae* strains that lacked, expressed or overexpressed Cup9p and/or Ubr1p were constructed and assayed for peptide import by growth on selective media. Suspensions of cells (auxotrophic for lysine) were serially diluted and plated on either rich media, minimal media lacking lysine and containing Lys-Ala dipeptide (selecting for peptide import), or minimal media containing the toxic dipeptide Leu-Eth (selecting against peptide import). All strains grew at comparable rates on rich media (Figure 3A). On minimal media that supplied the essential lysine as the Lys-Ala dipeptide, the *CUP9 UBR1* (wild-type), *cup9Δ UBR1* and *cup9Δ ubr1Δ* strains grew at comparable rates (Figure 3B), whereas the *CUP9 ubr1Δ* strain failed to grow (Figure 3B), in agreement with the observation that *UBR1* is required for peptide import (Figure 1). Opposite growth patterns were observed on media containing toxic dipeptide (Figure 3C). Comparison of the data in Figure 3 with *PTR2* mRNA levels (Figure 2A) suggested that *CUP9* exerts its effect on the import of peptides by repressing transcription of *PTR2*. For example, wild-type (*CUP9 UBR1*) cells overexpressing Cup9p exhibited reduced levels of *PTR2* mRNA (Figure 2A) and decreased sensitivity to toxic dipeptide (Figure 3C), whereas strains lacking *CUP9* (*cup9Δ UBR1* and *cup9Δ ubr1Δ*) overexpressed *PTR2* (Figure 2A) and were hypersensitive to toxic dipeptide (Figure 3C).

Cup9p is a repressor of the PTR2 gene

To address the mechanism of repression by Cup9p, we asked whether purified Cup9p (see Materials and methods) could selectively bind to specific regions of the *PTR2* promoter. Gel shift assays in the presence of poly-dI·dC competitor DNA were performed with labeled DNA

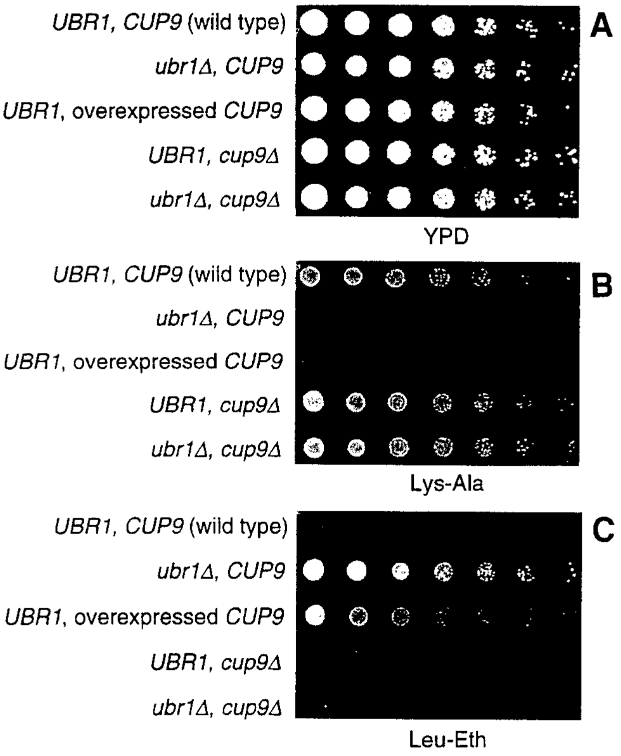


Fig. 3. Relative capacity for peptide import in congenic *S.cerevisiae* strains that contained, lacked or overexpressed Cup9p and/or Ubr1p. Serial dilutions of the indicated strains were deposited by a 48-pin applicator onto either rich (YPD) medium (**A**), minimal medium containing 66 μM Lys-Ala dipeptide as the sole source of lysine (**B**), or minimal medium containing both lysine (at 110 μM) and the toxic dipeptide Leu-Eth (at 55 μM) (**C**). The plates were incubated at 30°C for 1–2 days. The relevant genetic loci are shown on the left. The '*UBR1*, overexpressed *CUP9*' strain (JD52-2μ-CUP9) carried the high-copy plasmid pCB209 that expressed Cup9p from its natural promoter.

fragments of the *PTR2* promoter and His$_6$-tagged Cup9p (Cup9p-H$_6$) purified from *E.coli*. Cup9p-H$_6$ bound to a site within a distal region of the *PTR2* promoter (positions −448 to −897 relative to the inferred *CUP9* start codon), but did not bind to the proximal region of the *PTR2* promoter (positions −1 to −447) under the same conditions (Figure 2B).

The transcriptional repressor function of Cup9p was further suggested by the finding that the co-repressor complex Tup1p/Ssn6p plays a role in the control of peptide import. The Tup1p/Ssn6p complex inhibits transcription of many yeast genes through interactions with gene-specific DNA-binding repressors such as Matα2p (Chen *et al.*, 1993; Smith *et al.*, 1995), a homeodomain homolog of Cup9p. We found that most of our *sub* mutants that were not *CUP9* mutants could be complemented by low-copy plasmids bearing *SSN6* or *TUP1* (G.Turner, S.Saha and A.Varshavsky, unpublished data). In addition, deletion of *SSN6*, like deletion of *CUP9*, restored the ability of *ubr1Δ* cells to import peptides (data not shown).

Ubr1p-dependent degradation of Cup9p

The fact that deletion of *CUP9* renders *PTR2* transcription independent of Ubr1p (Figure 2A), and the observation that overexpression of Ubr1p enhances the import of peptides in Cup9p-expressing strains (data not shown) suggested that the N-end rule pathway regulates peptide import by targeting Cup9p for degradation. To test this conjecture, we carried out pulse–chase experiments with a C-terminally FLAG-tagged Cup9p in *UBR1* and *ubr1Δ* cells. Cup9p-FLAG was a very short-lived protein ($t_{1/2}$ ~5 min) in *UBR1* cells (Figure 4). By contrast, Cup9p was much longer-lived ($t_{1/2}$ >30 min) in *ubr1Δ* cells (Figure 4). Degradation of Cup9p was also found to depend upon *UBC2* and *UBC4* (data not shown), in agreement with the observation that a *ubc2Δ ubc4Δ* double mutant failed to import dipeptides (Figure 1G). The residual instability of Cup9p in *ubr1Δ* cells (Figure 4) suggested the presence of a second, Ubr1p-independent degron; this pattern is reminiscent of another homeodomain repressor, Matα2p, which also contains at least two distinct degradation signals (Hochstrasser and Varshavsky, 1990; Chen *et al.*, 1993).

Ubr1p recognizes engineered N-end rule substrates through their destabilizing N-terminal residues (Varshavsky, 1996). To determine the N-terminal residue of Cup9p, we overexpressed and purified Cup9p-FLAG from *ubr1Δ S.cerevisiae* and subjected it to N-terminal sequencing. Cup9p-FLAG was found to have a blocked (presumably acetylated) N-terminus (see Materials and methods), suggesting that Ubr1p targets Cup9p through a degron distinct from the N-degron. Independent evidence for this conjecture was produced through the analysis of GST-Cup9p-ha$_2$, a fusion of glutathione transferase (GST) and C-terminally ha-tagged Cup9p. Pulse–chase analysis of GST-Cup9p-ha$_2$ revealed that the fusion protein was nearly as short-lived *in vivo* as the N-terminally unmodified Cup9p-FLAG (Figure 4, lanes b–e; compare with Figure 5, lanes b–e). If Cup9p were targeted for processive degradation through a destabilizing N-terminal residue, a preliminary proteolytic cleavage(s) of Cup9p would be required to expose such a residue (Varshavsky, 1996). In the case of GST-Cup9p-ha$_2$, this cleavage would generate a proteolytic fragment consisting of GST and an N-terminal portion of Cup9p preceding the cleavage site. Since free GST has been found to be long-lived when expressed in yeast (data not shown), accumulation of such a proteolytic fragment would be expected to accompany the degradation of GST-Cup9p-ha$_2$.

Pulse–chase analysis of GST-Cup9p-ha$_2$, using glutathione–agarose beads to isolate GST-containing proteins (see Materials and methods), indicated that the degradation of GST-Cup9p-ha$_2$ by the N-end rule pathway was not accompanied by the appearance of a fragment containing the N-terminal GST moiety (Figure 5). This finding strongly suggested that Cup9p bears an 'internal' degron recognized by Ubr1p. Additional support for this conjecture was provided by truncation analysis of a Cup9p–DHFR–myc fusion protein. These experiments indicated that the N-terminal 81 residues of the 306-residue Cup9p are dispensable for its Ubr1p-dependent degradation, strongly suggesting that the relevant degradation signal resides in the C-terminal two-thirds of Cup9p (C.Byrd, I.Davydov and A.Varshavsky, unpublished data). Although these results are fully consistent with the presence of an internal Ubr1p-dependent degron in Cup9p, there remains the less parsimonious possibility that Cup9p is degraded via *trans*-targeting (Johnson *et al.*, 1990). In this process, the two determinants of the N-degron (a destabilizing N-terminal residue and a ubiquitin-accepting internal Lys

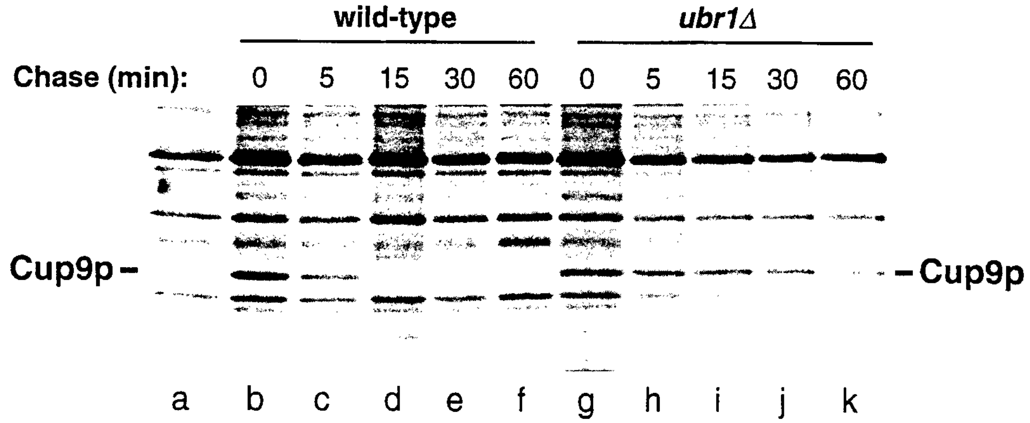


Fig. 4. Degradation of Cup9p by the N-end rule pathway. *Saccharomyces cerevisiae* strains carrying pCB210, a low-copy plasmid that expressed Cup9p-FLAG from the P_{CUP9} promoter, were labeled at 30°C with $[^{35}S]$methionine/cysteine for 5 min, followed by a chase for 0, 5, 15, 30 and 60 min, preparation of extracts, immunoprecipitation of Cup9p-FLAG with a monoclonal anti-FLAG antibody, SDS–PAGE and autoradiography/quantitation (see Materials and methods). Lane a, JD55 (*ubr1*Δ) cells that carried pCB200 (vector alone). Lanes b–f, JD52 (*UBR1*) (wild-type) cells that carried pCB210, expressing Cup9p-FLAG. Lanes g–k, same as lanes b–f, but with JD55 (*ubr1*Δ) cells.

residue) reside in two different polypeptides, whose interaction yields an active N-degron, leading to ubiquitinylation and degradation of the polypeptide bearing the Lys residue (Johnson *et al.*, 1990). At present, little is known about the protein ligands of Cup9p, save for the likely possibility that Cup9p interacts with the Tup1p/Ssn6p repressor complex (see above), similarly to the previously established interaction of this complex with Matα2, a homeodomain-containing homolog of Cup9p (Smith *et al.*, 1995).

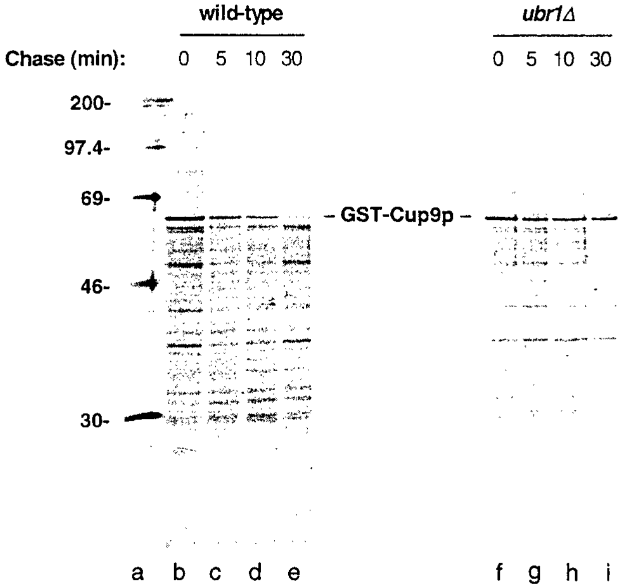


Fig. 5. Degradation of GST-Cup9p-ha_2 is not preceded by a processing that releases a GST-containing fragment. *Saccharomyces cerevisiae* strains carrying pCB120, a low-copy plasmid that expressed GST-Cup9p-ha_2 from the P_{GAL1} promoter, were labeled at 30°C with $[^{35}S]$methionine/cysteine for 5 min, followed by a chase for 0, 5, 10 and 30 min, preparation of extracts, isolation of GST-Cup9p-ha_2 on glutathione–agarose beads, SDS–PAGE and autoradiography/quantitation (see Materials and methods). Lanes b–e, JD52 (*UBR1*) (wild-type) cells expressing GST-Cup9p-ha_2. Lanes f–i, same as lanes b–e, but with JD55 (*ubr1*Δ) cells. Lane a, molecular mass markers (in kDa).

Discussion

The discovery that Ubr1p controls the import of peptides through degradation of the Cup9p repressor (Figure 6) identifies the first clear physiological function of the N-end rule pathway. The existence of mammalian, plant and bacterial homologs of the yeast Ptr2p transporter suggests that the import of peptides in these organisms may also be regulated by the N-end rule pathway.

Why was this pathway, rather than another Ub-dependent proteolytic system, recruited in the course of evolution for the regulation of peptide import? A plausible explanation is suggested by the ability of Ubr1p to bind peptides bearing destabilizing N-terminal residues (Varshavsky, 1996). Since more than half of the 20 amino acids are destabilizing in the yeast and mammalian N-end rules (Varshavsky, 1996), a significant fraction of imported peptides would be expected to compete with Cup9p for binding to Ubr1p. This competition may decrease the rate of Cup9p degradation. [Dipeptides added to a culture of *S.cerevisiae* have been shown to inhibit the N-end rule pathway (Baker and Varshavsky, 1991).] The ensuing increase in the level of Cup9p repressor would in turn decrease the production of the Ptr2p transporter, diminishing the rate of peptide import. This 'peptide-sensing' negative feedback loop would maintain the intracellular concentration of short peptides within a predetermined range—potentially a useful feature, made possible by the substrate-binding properties of Ubr1p. Experiments to verify this model are under way.

Another interesting aspect of the Ubr1p–Cup9p–Ptr2p regulatory circuit is its possible relevance to copper homeostasis. In addition to functioning as a repressor of peptide transport, Cup9p also contributes to the resistance of *S.cerevisiae* to copper toxicity during growth on lactate, a non-fermentable carbon source (Knight *et al.*, 1994). Copper homeostasis is mediated by a complex set of circuits, some of which also regulate iron metabolism (Zhou and Thiele, 1993; Ooi *et al.*, 1996). The double role of Cup9p in regulating both peptide import and copper homeostasis may signify a physiological connection between these seemingly unrelated processes.

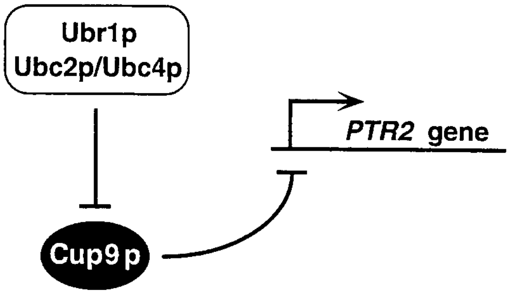


Fig. 6. A model for regulation of peptide import in *S.cerevisiae*. The expression of *PTR2*, which encodes a transmembrane peptide transporter, is regulated by the short-lived, homeodomain-containing transcriptional repressor Cup9p. The concentration of Cup9p is controlled in part through its degradation by the N-end rule pathway, whose targeting components include Ubr1p and either Ubc2p or Ubc4p (see the main text).

Our work adds Cup9p to the list of short-lived regulatory proteins whose degradation is mediated by the ubiquitin system. Many of these proteins are negative regulators. For example, Matα2p, a homeodomain-containing homolog of Cup9p, controls the mating type of *S.cerevisiae* (**a** or α) through repression of **a**-specific genes (Herskowitz, 1989). Matα2p appears to be constitutively short-lived in haploid cells (Hochstrasser and Varshavsky, 1990). Therefore, cessation of Matα2p synthesis upon the conversion of an α cell into an **a** cell results in rapid disappearance of Matα2p and the establishment of **a**-specific circuits (Hochstrassser, 1996). Progression of the cell cycle is also controlled by short-lived negative regulators, in particular by Sic1p, an inhibitor of CDK, the cyclin-dependent kinase (Schwob *et al.*, 1994). In this case, however, a rapid 'on-switch' is based on the phosphorylation-induced degradation of the previously stable Sic1p (King *et al.*, 1996). Whether the Cup9p repressor (Figure 6) is constitutively short-lived or whether its stability is regulated by external conditions such as, for example, different nitrogen sources, remains to be determined.

The finding that Cup9p lacks a destabilizing N-terminal residue indicates that Ubr1p, the recognition component of the N-end rule pathway, is able to target substrates bearing either internal degradation signals or N-degrons. The Gα subunit of the G protein—the other known physiological substrate of Ubr1p in *S.cerevisiae*—also lacks a destabilizing N-terminal residue (Madura and Varshavsky, 1994). Thus, a substantial fraction of naturally short-lived proteins targeted by the N-end rule pathway may bear internal degrons rather than N-degrons, a possibility that the identification of other physiological substrates of this pathway will address.

Materials and methods

Strains, media and genetic techniques

Saccharomyces cerevisiae strains were grown in rich (YPD) medium containing 2% peptone, 1% yeast extract and 2% glucose or in synthetic yeast media, containing 0.67% yeast nitrogen base without amino acids (Difco), auxotrophic nutrients at concentrations specified by Sherman *et al.* (1986) and either 2% glucose (SD medium), 2% raffinose (SM–raffinose medium) or 2% galactose (SM–galactose medium) as carbon sources. SHM–glucose medium used in halo assays was prepared according to Island *et al.* (1987) and was identical to SD medium except that it lacked methionine and contained allantoin (1 mg/ml) and yeast nitrogen base (Difco, 1.7 mg/ml) without amino acids and ammonium sulfate. *Escherichia coli* strain DH5α was used for plasmid propagation and cloning steps. For induction of the P_{GAL1} promoter, cells were grown to an A_{600} of 0.5–1 in SM–raffinose medium, pelleted and transferred to SM–galactose medium for 3 h.

Halo and dilution assays

Peptide import was assayed using the halo and dilution methods. For halo assays (Island *et al.*, 1987), cells were grown to an A_{600} of ~1 in SHM–glucose medium with auxotrophic supplements. Cells were pelleted by centrifugation and resuspended in water to $\sim 5\times10^7$ cells/ml. Then 0.1 ml of cell suspension was mixed with 10 ml of 0.8% noble agar (Difco) at 55°C, and spread on plates containing 20 ml of SHM medium. Sterile filter disks containing the toxic dipeptide L-leucyl-L-ethionine (Leu-Eth; 5 μmol) were placed in the middle of each plate, followed by incubation at 30°C for 1–2 days.

For dilution assays, strains were grown (under selection for plasmids) in SD medium to an A_{600} of ~1. Cells from each sample (1.5×10^7) were spun down and resuspended in 1 ml of water. The samples were serially diluted by 4-fold in a microtiter plate (150 μl/well) to generate eight different initial concentrations of cells that ranged from 1.5×10^7 to 915 cells/ml. The suspensions were spotted to various media, using a 48-pin applicator. The plates were incubated at 30°C for 1–2 days.

Leu-Eth was synthesized using standard methods of organic chemistry. Briefly, L-ethionine methyl ester (Eth-OMe) was produced from L-ethionine and methanol. Eth-OMe was then coupled with *N-tert*-butoxy-carbonyl-L-leucine-*p*-nitrophenyl ester (N-*t*-BOC-L-leucine-PNP), yielding N-*t*-BOC-L-leucyl-L-ethionine methyl ester, which was purified by flash chromatography on a silica column. N-*t*-BOC-L-leucyl-L-ethionine methyl ester was then converted into N-*t*-BOC-L-leucyl-L-ethionine by treatment with KOH. N-*t*-BOC-L-leucyl-L-ethionine was deprotected with trifluoroacetic acid, yielding Leu-Eth.

A screen for import-competent mutants in the ubr1Δ background

Thirty 5-ml cultures of JD83-1A (*leu2-3 ubr1Δ*), auxotrophic for leucine, were grown to an A_{600} of ~1, and $\sim1.5\times10^7$ cells from each culture were plated onto SD medium lacking leucine and histidine but containing the leucyl-histidine dipeptide at 0.23 μM. Plates were incubated at 30°C for 2 days, selecting for mutants able to grow on this medium. Of the 320 sub mutants ('suppressors of a block to peptide import in *ubr1Δ*') thus obtained, 199 were found by complementation tests to be clearly recessive. Among these, 101 mutants belonged to one complementation group, termed *sub1*.

Isolation of the CUP9 (SUB1) gene

A 50 ml culture of *S.cerevisiae* CBY15 (*ubr1Δ sub1-1*) was grown to an A_{600} of ~1 in SHM–glucose. Cells were made competent with lithium acetate (Ausubel *et al.*, 1992), and 25 0.1-ml samples ($\sim7\times10^7$ cells/ml) were transformed with a yeast genomic DNA library (American Type Culture Collection; #77164) in the *TRP1*, *CEN6*-based vector pRS200 (Sikorski and Hieter, 1989). The cells were pelleted, and each sample was resuspended in 1.5 ml of SHM–glucose and incubated at 30°C for 3 h. The cells were pelleted again, each sample was resuspended in 0.1 ml of water, added to 10 ml of 0.8% Noble agar at 55°C, and spread onto SHM–glucose plates lacking Trp and containing Leu-Eth at 37 μM. Of the $\sim2.5\times10^4$ Trp^+ transformants plated, 14 could grow in the presence of Leu-Eth. Of these, two yielded the library-derived plasmids, pSUB1-1 and pSUB1-6, that complemented the *sub1* mutation of CBY15. The insert of pSUB1-6 contained the insert of pSUB1-1. Partial sequencing of the ~8.2 kb insert of pSUB1-1 identified it as a region of the *S.cerevisiae* chromosome XVI. Deletion analysis (not shown) localized the complementing activity to an ~2.8 kb *Hin*dIII–*Kpn*I fragment, whose sequence revealed the presence of two ORFs (*CUP9* and SCYPL178W). Further deletion analysis (not shown) localized the complementing activity to the previously isolated (Knight *et al.*, 1994) *CUP9* gene. Verification that *CUP9* and *SUB1* were one and there same gene was carried out as described in Results.

Construction of null cup9 mutants

The ~2.8 kb *Hin*dIII–*Kpn*I genomic DNA fragment containing *CUP9* was ligated to *Hin*dIII–*Kpn*I-cut pRS306ΔSpeI, yielding pCB117. pRS306ΔSpeI was derived from the *URA3*-bearing pRS306 (Sikorski and Hieter, 1989) through elimination of its *Spe*I site. The *CUP9* ORF was disrupted by inserting an ~2 kb, blunt-ended, *LEU2*-containing *Sal*I fragment from pJJ283 into the *Spe*I site of *CUP9*, yielding pCB119, in which the *LEU2* and *CUP9* ORFs were oriented in opposite directions. An ~5 kb *Hin*dIII–*Pvu*II fragment of pCB119 containing the *cup9::LEU2*

Table I. *Saccharomyces cerevisiae* strains used in this study

Strain	Genotype	References
DF5	*MAT*a/*MATα trp1-1/trp1-1 ura3-52/ura3-52 his3-Δ200/his3-Δ200 leu2-3,112/leu2-3,112 lys2-801/lys2-801 gal/gal*	Finley *et al.* (1987)
EJ1	*MATα ubc5Δ::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal*	Derivative of DF5[a]
EJ2	*MATα ubc1Δ::URA3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal*	Derivative of DF5[a]
EJ3	*MATα ubc4Δ::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal*	Derivative of DF5[a]
EJ10	*MATα ubc4Δ::HIS3 ubc5Δ::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal*	Derivative of DF5[a]
EJY102	*MATα ubc2Δ::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal*	Derivative of DF5[a]
EJY105	*MAT*a *ubc2Δ::LEU2 ubc4Δ::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal*	Derivative of DF5[a]
EJY106	*MATα ubc2Δ::LEU2 ubc4Δ::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal*	Derivative of DF5[a]
MHY495	*MATα ubc6-Δ1::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of DF5[b]
MHY503	*MATα ubc4-Δ1::HIS3 ubc6-Δ1::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of DF5[b]
MHY507	*MATα ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of DF5[b]
MHY550	*MATα ubc4-Δ2::TRP1 ubc5-Δ1::LEU2 ubc6-Δ1::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of DF5[b]
MHY552	*MATα ubc6-Δ1::HIS3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of DF5[b]
MHY554	*MATα ubc4-Δ1::HIS3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of DF5[b]
MHY557	MATα ubc4-Δ1::HIS3 ubc6-Δ1::HIS3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801	Derivative of DF5[b]
MHY570	*MATα ubc4-Δ2::TRP1 ubc5-Δ1::LEU2 ubc6-Δ1::HIS3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of DF5[b]
MHY599	*MATα pas2 trp ura3-52 ade1 leu2-3*	Chen *et al.* (1993)
MHY601	*MATα ubc8::URA3 trp1-1 ura3-1 ade2-1 his3-11 leu2-3,112 can1-100*	Chen *et al.* (1993)
YWO55	*MATα ubc9-1 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of DF5[c]
YPH500	*MATα trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-Δ1 lys2-801*	Sikorski and Hieter (1989)
BBY67	*MATα ubc2Δ::LEU2 trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-Δ1 lys2-801*	Derivative of YPH500[d]
KM207-1	*MATα ate1Δ::TRP1 trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-Δ1 lys2-801*	Derivative of YPH500[e]
JD34	*MAT*a *ubc2Δ::URA3 trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-Δ1 lys2-801*	Derivative of YPH500[f]
JD51	*MAT*a/*MATα trp1-Δ63/trp1-Δ63 ura3-52/ura3-52 his3-Δ200/his3-Δ200 leu2-3,112/leu2-3,112 lys2-801/lys2-801*	Dohmen *et al.* (1995)
JD52	*MAT*a *trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Johnson *et al.* (1995)
JD53	*MATα trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Dohmen *et al.* (1995)
JD55	*MAT*a *ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Madura and Varshavsky (1994)
JD83-1A	*MATα ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of JD51[f]
RJD549	*MAT*a *cdc34-1 trp1 ura3-52 leu2-3*	R.Deshaies[g]
RJD795	*MAT*a *cdc34-2 trp1 ura3-52 leu2-3*	R.Deshaies[g]
SGY6	*MATα nta1Δ::TRP1 trp1-Δ63 ura3-52 ade2-101 his3-Δ200 leu2-3,112 lys2-801*	S.Grigoryev[h]
CBY4	*MATα ubc2Δ::URA3 ubc5Δ::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal*	Derivative of EJ1
CBY5	*MATα ubc2Δ::LEU2 ubc6-Δ1::HIS3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of MHY495
CBY6	*MATα ubc2Δ::URA3 ubc7::LEU2 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of MHY507
CBY7	*MATα ubc2Δ::LEU2 ubc8::URA3 trp1-1 ura3-1 ade2-1 his3-11 leu2-3,112 can1-100*	Derivative of MHY601
CBY8	*MATα ubc2Δ::LEU2 ubc1Δ::URA3 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801 gal*	Derivative of EJ2
CBY9	*MATα ubc2Δ::LEU2 ubc9-1 trp1-1 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of YWO55
CBY10	*MATα ubc2Δ::LEU2 pas2 trp ura3-52 ade1 leu2-3*	Derivative of MHY599
CBY11	*MAT*a *ubc2Δ::LEU2 cdc34-1 trp1 ura3-52 leu2-3*	Derivative of RJD549
CBY15	*MATα sub1-1 ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of JD83-1A
CBY16	*MAT*a *cup9::LEU2 ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of JD55
CBY17	*MAT*a *cup9::LEU2 ubr1Δ::HIS3 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of JD55
CBY19	*MAT*a *cup9::LEU2 trp1-Δ63 ura3-52 his3-Δ200 leu2-3,112 lys2-801*	Derivative of JD52
CBY23	*MAT*a/*MATα sub1-1/cup9Δ::LEU2 ubr1Δ::HIS3/ubr1Δ::HIS3 trp1-Δ63/trp1-Δ63 ura3-52/ura3-52 his3-Δ200/his3-Δ200 leu2-3,112/leu2-3,112 lys2-801/lys2-801*	Produced by mating CBY15 and CBY16
CBY24	*MAT*a/*MATα sub1-1/cup9Δ::LEU2 ubr1Δ::HIS3/ubr1Δ::HIS3 trp1-Δ63/trp1-Δ63 ura3-52/ura3-52 his3-Δ200/his3-Δ200 leu2-3,112/leu2-3,112 lys2-801/lys2-801*	Produced by mating CBY15 and CBY17

[a] Johnson *et al.* (1992, 1995). A gift from E.Johnson, the Rockefeller University, New York, NY 10021-6399, USA.

[b] Chen *et al.* (1993). A gift from M.Hochstrasser, Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637, USA.

[c] A gift from S.Jentsch, ZMBH, Universität Heidelberg, 69120 Heidelberg, Germany.

[d] Dohmen *et al.* (1990).

[e] A gift from K.Madura, Department of Biochemistry, UMDNJ-Johnson Medical School, Piscataway, NJ 08854, USA.

[f] A gift from J.Dohmen, Heinrich-Heine-Universität, Institut für Mikrobiologie, 40225 Düsseldorf, Germany.

[g] A gift from R.Deshaies, Division of Biology, Caltech, Pasadena, CA 91125, USA.

[h] A gift from S.Grigoryev, Department of Biology, University of Massachusetts, Amherst, MA 01003, USA.

disruption allele was used to replace the wild-type *CUP9* alleles of JD52 (wild-type) and JD55 (*ubr1Δ*) by homologous recombination (Rothstein, 1991), generating strains CBY19 and CBY17, respectively.

CUP9-expressing plasmids

The plasmid pCB116, which expressed Cup9p-FLAG from the P_{GAL1} promoter, was constructed by subcloning an ~1 kb *Bam*HI–*Eco*RI fragment containing the *CUP9-FLAG* ORF into the *Bam*HI–*Eco*RI site(s) of p416GAL1 (Mumberg *et al.*, 1994). The *CUP9-FLAG*-containing fragment was constructed by PCR amplification of the *CUP9* ORF of plasmid pCB111 using primers PCB1 (5′-CGCGGATCCGAATAGTTACATTCGAAGATG-3′) and PCB6 (5′-CCGGAATTCTCATTTATCATCATCGTCTTTGTAATCATTCATATCAGGGTTGGATAG-3′), resulting in the addition of the 8-residue FLAG epitope, DYKDDDDK, to the C-terminus of Cup9p. pCB111 was constructed by subcloning the ~2.8 kb *Hin*dIII–*Kpn*I fragment of the *CUP9*-containing pSUB1-1 (see above) into *Hin*dIII–*Kpn*I-cut pRS316 (Sikorski and Hieter, 1989). pCB202 was constructed by subcloning an ~1 kb fragment containing the P_{CUP9} promoter into *Hin*dIII–*Bam*HI-cut pCB201. [The ~1 kb fragment was produced by PCR from pCB111 using primers PCB8

(5′-GTGTTAGTAAGCTTGTAAAGGAATGCACGTATT-3′) and PCB9 (5′-CCCGCGGATCCGCATGCAACTATTCTCGAAGGTTGT-3′).] pCB200 (*ARS-CEN*, *LEU2*) and pCB201 (2μ, *LEU2*) were constructed by replacing the 517 bp *Sca*I–*Eco*RI fragment of pBR322 (Ausubel *et al.*, 1992) with, respectively, the 3822 bp *Sca*I–*Nae*I fragment of pRS415 (Sikorski and Hieter, 1989) and the 4650 bp *Sca*I–*Nae*I fragment of pRS425 (Christianson *et al.*, 1992).

The plasmid pCB209 (2μ, *LEU2*), which expressed *CUP9* from the P_{CUP9} promoter, was constructed by replacing the *Sph*I–*Sal*I fragment of pCB202 with an ~1 kb fragment containing the *CUP9* ORF that was produced by PCR from pCB111, using primers PCB10 (5′-CCCGCGGATCCGCATGCGAAGATGAATTATAACTGC-3′) and PCB12 (5′-CCCGCGCGGTCGACCTCAATTCATATCAGGGTTGGATAG-3′). pCB210 (*ARS-CEN*, *LEU2*) that expressed Cup9p-FLAG from the P_{CUP9} promoter was constructed by replacing the *Sph*I–*Sal*I fragment of pCB202 with an ~1 kb fragment containing the *CUP9-FLAG* ORF, which was produced from pCB111 using primers PCB10 (see above) and PCB13 (5′-CCCGCGCGGTCGACCTCATTTATCATCATCGTCTTTGTAATCATTCATATCAGGGTTGGATAG-3′), yielding pCB211. The ~2 kb *Hin*dIII–*Sal*I fragment of pCB211 containing P_{CUP9} and the *CUP9-FLAG* ORF was subcloned into pCB200, yielding pCB210.

Plasmid pCB120, expressing GST-Cup9p-ha$_2$ from the P_{GAL1} promoter, was constructed by subcloning the ~1.6 kb *Xba*I–*Eco*RI fragment, containing the *GST-CUP9-ha$_2$* ORF, into the *Xba*I–*Eco*RI site(s) of p416GAL1. The *Xba*I–*Eco*RI fragment was produced by PCR amplification of the *GST-CUP9-ha$_2$* ORF of pGEX-2T–CUP9-ha$_2$, using the primers PCB3 (5′-CCGGAATTCTCAAGCGTAATCTGGAACATCGTATGGGTAAGCGTAATCTGGAACATCGTATGGGTAATTCATATCAGGGTTGGATAG-3′) and PCB5 (5′-TGCTCTAGAACAGTATTCATGTCCCCTATA-3′). pGEX-2T–CUP9-ha$_2$ was constructed by subcloning an ~1 kb fragment containing the *CUP9-ha$_2$* ORF into the *Bam*HI–*Eco*RI site(s) of pGEX-2T (Pharmacia), resulting in an in-frame fusion of the sequence encoding 26 kDa glutathione *S*-transferase (GST) domain of *Saccharomyces japonicum* (Smith and Johnson, 1988) to the second codon of *CUP9*. The *CUP9-ha$_2$*-containing fragment was produced by PCR amplification of the *CUP9* ORF of pCB111 using the primers PCB3 (see above) and PCB4 (5′-CGCGGATCCAATTATAACTGCGAAATACAAAAC-3′). This step added to the C-terminus of Cup9p a sequence encoding a tandem repeat of the 9-residue sequence YPYDVPDYA, derived from hemagglutinin (ha) of influenza virus.

Northern hybridization

RNA was isolated from *S.cerevisiae* as described (Schmitt *et al.*, 1990). Electrophoresis of the RNA samples was carried out on a formaldehyde RNA gel (Ausubel *et al.*, 1992). An ~50-μg RNA sample was loaded on a 1% agarose gel containing 1× MOPS buffer, 0.74% (v/v) formaldehyde, 1.9 mg/ml iodoacetamide and 0.5 μg/ml ethidium bromide. Electrophoresis was carried out in 1× MOPS buffer at 5 V/cm. RNA was transferred to BrightStar-Plus membrane (Ambion) using TurboBlotter (Schleicher & Schuell) and Ambion RNA transfer buffer. RNA was crosslinked to the air-dried membranes using 254 nm light (Ausubel *et al.*, 1992).

DNA probes were prepared by the random priming method (Ausubel *et al.*, 1992) using [^{32}P]dCTP and a DNA labeling kit (Pharmacia). Hybridization was carried out for 8–16 h at 42°C in Prehybridization/Hybridization Solution (Ambion). Filters were washed according to the manufacturer's protocol and subjected to autoradiography.

Gel shift assay

PCR was used to extend the Cup9p ORF with a sequence encoding Ser-Gly-Gly-Thr-His$_6$, yielding Cup9p-H$_6$, and to engineer flanking restriction sites (*Nde*I and *Bam*HI) for insertion into pET-11c (Novagen). Cup9p-H$_6$ was overexpressed in *E.coli* BL21 (DE3) (Novagen) (Ausubel *et al.*, 1992) and purified on a 3 ml Ni-NTA column (Qiagen), using a linear gradient of imidazole. Cup9p-H$_6$ eluted at ~0.25 M imidazole (~90% pure at this step); it was dialyzed at 4°C against 10% glycerol, 0.1 M KCl, 1 mM EDTA, 0.5 mM dithiothreitol, 20 mM HEPES, pH 7.9, and then snap-frozen in multiple samples in liquid N$_2$, and stored at –80°C. The proximal (–1 to –447) and distal (–448 to –897) *PTR2* promoter probes for the gel shift assay were constructed by PCR amplification in the presence of [α-^{32}P]dCTP, and were purified using spin columns (Qiagen). The gel shift reactions (20 μl) contained 50 μg/ml poly-dI·dC (Pharmacia); ~1.5 μg/ml (500 c.p.m.) DNA probe; 1 mg/ml acetylated serum albumin (New England Biolabs) and either 0.1, 0.5 or 1 μg/ml of Cup9p-H$_6$ in 10% glycerol, 0.1 M KCl, 2.5 mM MgCl$_2$, 1 mM EDTA, 20 mM HEPES, pH 7.9. The samples were incubated for 30 min at room temperature, then loaded onto a 4% polyacrylamide gel (40:1, acrylamide:*bis*-acrylamide) in 0.5× TBE (Ausubel *et al.*, 1992), and electrophoresed at 10 V/cm for 3 h at 4°C, followed by autoradiography.

Pulse–chase analysis of Cup9p

One hundred ml cultures of *S.cerevisiae* JD52 (*UBR1*) and JD55 (*ubr1*Δ) carrying either pCB210 (expressing Cup9p-FLAG from the *CUP9* promoter) or pCB200 (vector alone) were grown to an A_{600} of ~1 in SD(–Leu) medium. Cells (50 A_{600} units total) from each of the four cultures were gently pelleted by centrifugation, washed with 5 ml of SD(–Leu), pelleted again, resuspended in 2 ml of SD(–Leu), and incubated at 30°C for 10 min. Each sample was labeled for 5 min with 1.4 mCi of [^{35}S]methionine/cysteine (EXPRESS, New England Nuclear) at 30°C, followed by pelleting in a microfuge for ~15 s. The cells were resuspended in 2.6 ml of SD(–Leu), 5 mM L-methionine, 5 mM L-cysteine, and incubated at 30°C. Samples of 0.5 ml were withdrawn during the incubation, pelleted and resuspended in 0.15 ml of 0.5 M NaCl-Lysis Buffer (1% Triton X-100, 0.5 M NaCl, 5 mM EDTA, 50 mM Na-HEPES, pH 7.5) containing a mixture of protease inhibitors (Ghislain *et al.*, 1996). Glass beads (0.5 mm) were added, and cells were disrupted by vortexing (six times, for 30 s each, with 1 min incubations on ice in between), followed by the adjustment of NaCl concentration to 0.15 M through the addition of 75 mM NaCl-Lysis Buffer, further vortexing for 30 s, and centrifugation at 12 000 *g* for 10 min. The volumes of supernatants were adjusted to equalize the amounts of 10% trichloroacetic acid-insoluble ^{35}S. Cup9p-FLAG was immunoprecipitated by the addition of 20 μl of the monoclonal anti-FLAG M2 antibody conjugated to agarose beads (Kodak). Suspensions were incubated at 0°C for 1 h, with rotation, then centrifuged at 12 000 *g* for 30 s, and washed four times with 0.8 ml of 0.15 M NaCl-Lysis Buffer. The pellets were resuspended in SDS-sample buffer, heated at 100°C for 3 min, and subjected to SDS–12% PAGE, followed by autoradiography and quantitation using a PhosphorImager (Molecular Dynamics).

Pulse–chase analysis of GST-Cup9p-ha$_2$ was carried out as described by Bartel *et al.* (1990). Approximately 10 A_{600} units of galactose-induced cells were labeled for 5 min with 0.3 mCi of [^{35}S]EXPRESS in 400 μl SM–galactose (–Ura) at 30°C. The cells were then transferred to microfuge tubes, pelleted and resuspended in 500 μl of SD (–Ura), 5 mM L-methionine, 5 mM L-cysteine. Samples of 0.1 ml were withdrawn during the incubation, pelleted and lysed as above. The ^{35}S-labeled GST-Cup9p-ha$_2$ was purified using glutathione–agarose beads (Sigma) which had been blocked with bovine serum albumin (BSA; 10 mg/ml). Twenty μl of glutathione–agarose beads were added to each sample and the suspensions were incubated at 0°C for 60 min, with rotation, followed by washes and electrophoretic analyses as described for Cup9p-FLAG.

Purification and N-terminal sequencing of Cup9p-FLAG

Four 2-l cultures of JD55 (*ubr1*Δ) carrying pCB116 that expressed Cup9p-FLAG from the *GAL1* promoter were grown under selection in SM–raffinose to an A_{600} of ~0.8, followed by transfer to SM–galactose and incubation at 30°C for 3 h. Longer induction times resulted in a Cup9p-mediated cytotoxicity and lower yields of Cup9p-FLAG. The cells (~1×10^{11}) were harvested and lysed at 4°C as described by Burgers (1995). The extract was fractionated by precipitation with 0.4% Polymin P (Sigma) and then further by precipitation with 48% saturated ammonium sulfate. The pellet was dissolved in 3 ml of TBS buffer (0.15 M NaCl, 50 mM Tris–HCl, pH 7.5), and passed through Sephadex G-25 in TBS. The resulting sample (8.3 ml, ~70 mg/ml of protein) was applied to a column (1 ml) of the monoclonal anti-FLAG M2 antibody (Kodak). The column was washed three times with 10 ml of TBS, and Cup9p-FLAG was eluted by the addition of five 1-ml samples of TBS containing, respectively, 50, 100, 100, 200 and 200 μg/ml of the FLAG peptide (Kodak). Peak Cup9p-FLAG fractions (detected by immunoblotting) were concentrated by partial lyophilization, followed by precipitation with methanol (Wessel and Flügge, 1984) in the presence of human insulin (Sigma, 0.3 mg/ml) as a carrier. The resulting sample was fractionated by SDS–12% PAGE and electroblotted onto Pro-Blot membrane (Perkin-Elmer). After a brief staining with Coomassie, the band of the 37 kDa Cup9p-FLAG (~15 pmol) was excised from the membrane. Half of the sample was used to determine the amino acid composition; the other half was subjected to N-terminal sequencing for seven cycles, using the Applied Biosystems 476A protein sequencer at the Caltech Microchemistry Facility.

Acknowledgements

We thank R.Deshaies, R.J.Dohmen, S.Grigoryev, M.Hochstrasser, S.Jentsch, E.S.Johnson, K.Madura, I.Ota and C.Trotta for the gifts of strains and plasmids; S.Carter for his guidance in the synthesis of Leu-Eth; N.Johnsson, R.Deshaies, R.J.Dohmen and A.Webster for helpful discussions; L.Peck, Y.T.Kwon, A.Webster and F.Du for comments on the manuscript; S.Saha for help in overexpressing Cup9p; and N.Riley for technical assistance. This work was supported by grants to A.V. from the National Institutes of Health (DK39520 and GM31530).

References

Alagramam,K., Naider,F. and Becker,J.M. (1995) A recognition component of the ubiquitin system is required for peptide transport in *Saccharomyces cerevisiae*. *Mol. Microbiol.*, **15**, 225–234.

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Smith,J.A., Seidman,J.G. and Struhl,K. (1992) *Current Protocols in Molecular Biology*. Wiley-Interscience, New York.

Bachmair,A. and Varshavsky,A. (1989) The degradation signal in a short-lived protein. *Cell*, **56**, 1019–1032.

Bachmair,A., Finley,D. and Varshavsky,A. (1986) *In vivo* half-life of a protein is a function of its amino-terminal residue. *Science*, **234**, 179–186.

Baker,R.T. and Varshavsky,A. (1991) Inhibition of the N-end rule pathway in living cells. *Proc. Natl Acad. Sci. USA*, **87**, 2374–2378.

Baker,R.T. and Varshavsky,A. (1995) Yeast N-terminal amidase: a new enzyme and component of the N-end rule pathway. *J. Biol. Chem.*, **270**, 12065–12074.

Balzi,E., Choder,M., Chen,W.A., Varshavsky,A. and Goffeau,A. (1990) Cloning and functional analysis of the arginyl-tRNA–protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **265**, 7464–7471.

Bartel,B. (1990) *Molecular Genetics of the Ubiquitin System: the Ubiquitin Fusion Proteins and Proteolytic Targeting Mechanisms*. PhD Thesis, M.I.T., Cambridge, MA, USA.

Bartel,B., Wünning,I. and Varshavsky,A. (1990) The recognition component of the N-end rule pathway. *EMBO J.*, **9**, 3179–3189.

Burgers,P.M.J. (1995) Preparation of extracts from yeast and avoidance of proteolysis. *Methods Mol. Cell. Biol.*, **5**, 330–335.

Chau,V., Tobias,J.W., Bachmair,A., Marriott,D., Ecker,D.J., Gonda,D.K. and Varshavsky,A. (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science*, **243**, 1576–1583.

Chen,P., Johnson,P., Sommer,T., Jentsch,S. and Hochstrasser,M. (1993) Multiple ubiquitin-conjugating enzymes participate in the *in vivo* degradation of the yeast MATα2 repressor. *Cell*, **74**, 357–369.

Christianson,T.W., Sikorski,R.S., Dante,M. and Hieter,P. (1992) Multifunctional yeast high-copy-number shuttle vectors. *Gene*, **110**, 119–122.

deGroot,R.J., Rümenapf,T., Kuhn,R.J. and Strauss,J.H. (1991) Sindbis virus RNA polymerase is degraded by the N-end rule pathway. *Proc. Natl Acad. Sci. USA*, **88**, 8967–8971.

Dohmen,R.J., Stappen,R., McGrath,J.P., Forrová,H., Kolarov,J., Goffeau,A. and Varshavsky,A. (1995) An essential yeast gene encoding a homolog of ubiquitin-activating enzyme. *J. Biol. Chem.*, **270**, 18099–18109.

Finley,D., Özkaynak,E. and Varshavsky,A. (1987) The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation and other stresses. *Cell*, **48**, 1035–1046.

Ghislain,M., Dohmen,R.J., Lévy,F. and Varshavsky,A. (1996) Cdc48p interacts with Ufd3p, a WD-repeat protein required for ubiquitin-dependent proteolysis in *Saccharomyces cerevisiae*. *EMBO J.*, **15**, 4884–4899.

Herskowitz,I. (1989) A regulatory hierarchy for cell specialization in yeast. *Nature*, **342**, 749–757.

Hilt,W. and Wolf,D.H. (1996) Proteasomes: destruction as a programme. *Trends Biochem. Sci.*, **21**, 96–102.

Hochstrasser,M. (1996) Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.*, **30**, 405–439.

Hochstrasser,M. and Varshavsky,A. (1990) *In vivo* degradation of a transcriptional regulator: the yeast α2 repressor. *Cell*, **61**, 697–708.

Island,M.D., Naider,F. and Becker,J.M. (1987) Regulation of dipeptide transport in *S.cerevisiae* by micromolar amino acid concentrations. *J. Bacteriol.*, **169**, 2132–2136.

Jentsch,S. (1992) The ubiquitin-conjugating system. *Annu. Rev. Genet.*, **26**, 179–207.

Jentsch,S. and Schlenker,S. (1995) Selective protein degradation: a journey's end within the proteasome. *Cell*, **82**, 881–884.

Johnson,E.S., Gonda,D.K. and Varshavsky,A. (1990) *Cis–trans* recognition and subunit-specific degradation of short-lived proteins. *Nature*, **346**, 287–291.

Johnson,E.S., Bartel,B., Seufert,W. and Varshavsky,A. (1992) Ubiquitin as a degradation signal. *EMBO J.*, **11**, 497–505.

Johnson,E.S., Ma,P.C.M., Ota,I.M. and Varshavsky,A. (1995) A proteolytic pathway that recognizes ubiquitin as a degradation signal. *J. Biol. Chem.*, **270**, 17442–17456.

King,R.W., Deshaies,R.J., Peters,J.M. and Kirschner,M.W. (1996) How proteolysis drives the cell cycle. *Science*, **274**, 1652–1659.

Knight,S.A.B., Tamai,K.T., Kosman,D.J. and Thiele,D.J. (1994) Identification and analysis of a *Saccharomyces cerevisiae* copper homeostasis gene encoding a homeodomain protein. *Mol. Cell. Biol.*, **14**, 7792–7804.

Madura,K. and Varshavsky,A. (1994) Degradation of Gα by the N-end rule pathway. *Science*, **265**, 1454–1458.

Madura,K., Dohmen,R.J. and Varshavsky,A. (1993) N-recognin/Ubc2 interactions in the N-end rule pathway. *J. Biol. Chem.*, **268**, 12046–12054.

Mumberg,G., Müller,R. and Funk,M. (1994) Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Res.*, **22**, 5767–5768.

Ooi,C.E., Rabinovich,E., Dancis,A., Bonifacino,J.S. and Klausner,R.D. (1996) Copper-dependent degradation of the *Saccharomyces cerevisiae* plasma membrane copper transporter Ctr1p in the apparent absence of endocytosis. *EMBO J.*, **15**, 3515–3523.

Rothstein,R. (1991) Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods Enzymol.*, **194**, 281–301.

Rubin,D.M., van Nocker,S., Glickman,M., Coux,O., Wefes,I., Sadis,S., Fu,H., Goldberg,A., Vierstra,R. and Finley,D. (1997) ATPase and ubiquitin-binding proteins of the yeast proteasome. *Mol. Biol. Rep.*, **24**, 17–26.

Schmitt,M.E., Brown,T.A. and Trumpower,B.L. (1990) A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **18**, 3091–3092.

Schwob,E., Bohm,T., Mendenhall,M.D. and Nasmyth,K. (1994) The B-type cyclin kinase inhibitor p40 (Sic1) controls the G1 to S transition in *Saccharomyces cerevisiae*. *Cell*, **79**, 233–244.

Sherman,F., Fink,G.R. and Hicks,J.B. (1986) *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sikorski,R.S. and Hieter,P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *S. cerevisiae*. *Genetics*, **122**, 19–27.

Smith,D.B. and Johnson,K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene*, **67**, 31–38.

Smith,R.L., Redd,M.J. and Johnson,A.D. (1995) The tetratricopeptide repeats of Ssn6 interact with the homeodomain of α2. *Genes Dev.*, **9**, 2903–2910.

Varshavsky,A. (1996) The N-end rule: functions, mysteries, uses. *Proc. Natl Acad. Sci. USA*, **93**, 12142–12149.

Wessel,D. and Flügge,U.I. (1984) A method for quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.*, **138**, 141–143.

Wolberger,C. (1996) Homeodomain interactions. *Curr. Opin. Struct. Biol.*, **6**, 62–68.

Zhou,P. and Thiele,D.J. (1993) Copper and gene regulation in yeast. *Biofactors*, **4**, 105–115.

Received August 13, 1997; revised October 13, 1997; accepted October 14, 1997